SEP 23 1922
UNIV. OF MICH.

Journal of Educational Research

OFFICIAL ORGAN OF THE
EDUCATIONAL RESEARCH ASSOCIATION

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
FRANK N. FREEMAN

VOL. VI SEPTEMBER, 1922 No. 2

CONTENTS

		Page
ANOTHER EDUCATIONAL CAMPAIGN	J. W. Richardson	97
RECENT DEVELOPMENTS IN SILENT READING TESTS	C. R. Stone	102
SOME ESSENTIALS IN SCHOOL SUPERVISION	E. J. Ashbaugh	116
CAN TEACHERS SELECT BRIGHT AND DULL PUPILS?	G. F. Verner	126
THE MENTAL AGE OF ADULTS	E. A. Lincoln	133
ATTEMPTS AT TEST VALIDATION	Raymond Franzen	145
EDITORIALS		159
REVIEWS AND ABSTRACTS		162
NEWS ITEMS AND COMMUNICATIONS		172
EDUCATIONAL RESEARCH ASSOCIATION		180

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1922, by
Public School Publishing Co.

Official Organ of the Educational Research Association

Editor-in-Chief, B. R. Buckingham
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ohio State University, Columbus, Ohio.*

W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

FRANK N. FREEMAN, *Ex-Officio, Secretary Educational Research Association.*

GEORGE D. STRAYER, *Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California*

GUY M. WHIPPLE, *University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Nebr.*

M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota, Minneapolis.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin, Madison.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean of University Administration, University of Kansas, Lawrence, Kans.*

ROBT. H. LANE, *Assistant Superintendent of Schools Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.*

LEON O. SMITH, *Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.*

W. W. THEISEN, *Director, Division of Reference and Research, Public Schools, Cleveland, Ohio.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Mich.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Blooming[illegible] monthly, except July and August by the Public School Publishing Co. The subscription [illegible] $4.20 in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on edito[illegible] B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume VI SEPTEMBER, 1922 Number 2

ANOTHER EDUCATIONAL CAMPAIGN

J. W. RICHARDSON
Director, Bureau of Research, Hibbing, Minnesota

In a former account of the use of campaign principles in elementary-school work, called "The Campaign Method in Elementary Education,"[1] we limited our discussion to progress in arithmetic under certain conditions. The conditions chosen were those which we recognize by the collective term of "a campaign." As stated in the above mentioned report, "A campaign is characterized: first, by a recognized need or desire on the part of those who organize it; second, by definite time limits, particularly for the close; third, by a definitely organized effort to stimulate and sustain interest. A fourth characteristic rests upon the fact that the success or failure of a campaign is determined by individuals other than those who organize it, whether it be by voters in the political compaign, converts in the religious revival, or subscribers to the fund. Furthermore, the compaign has a peculiar value in awakening the individual to the duties and privileges of co-operation whether civic, social, moral, or intellectual."

In making that report, we stated the belief that the plan would be just as feasible in the formal or mechanical phases of any elementary school work as in the four fundamental processes in whole numbers. We are now ready to state that the plan is certainly just as feasible in the subject of spelling as in arithmetic.

[1]Reported in the *Journal of Educational Research*, 2:481, June, 1920.

The Spelling Campaign

The Pupils—Twenty classes in fourth, fifth, and sixth grades, in two buildings, were each assigned to one of two groups. One of these groups was to be campaigned, and was designated as the experimental group; the other group, which was to take the initial and final tests, but was not to be campaigned, was called the control group. Each group consisted of ten classes, and both groups were estimated by principals to be of approximately equal ability.

The Material—Ten lists of fifty words each, with illustrative phrases, were used for tests. Each list was made up of the first fifty words in columns H to Q inclusive of the Ayres' Scale.

The Plan—Teachers were furnished a copy of the day's list on alternate mornings. They gave the test by reading the illustrative prase containing each word, and then repeating the word to be written in list form by the pupils.

The papers from the ten classes in the experimental group were corrected the day they were written and individual lists of misspelled words were returned to the teacher to be given to the pupils for study in the spelling period the next day. The pupils of this group were also told that the test would be repeated in exactly a month. No lists were returned nor announcements made to the classes in the control group.

TABLE I. PERCENT OF CORRECT SPELLINGS GRADE IV

Ayres' List	Standard	Experimental Group		Control Group	
		First	Second	First	Second
H	98	96	98	95	96
I	96	96	99	91	93
J	94	96	98	92	94
K	92	94	97	84	86
L	88	90	93	81	84
M	84	83	94	73	78
N	79	81	93	67	71
O	73	73	91	58	66
P	66	63	88	46	44
Q	58	58	79	44	47
Total		84	93	73	76

TABLE II. PERCENT OF CORRECT SPELLINGS GRADE V

Ayres' List	Standard	Experimental Group		Control Group	
		First	Second	First	Second
H	100	99	99	98	98
I	99	97	98	95	97
J	98	98	99	96	97
K	96	97	98	95	95
L	94	94	96	92	93
M	92	91	95	86	88
N	88	88	92	85	87
O	84	87	94	82	83
P	79	78	92	70	69
Q	73	75	89	65	68
Total		90	95	86	87

TABLE III. PERCENT OF CORRECT SPELLINGS GRADE VI

Ayres' List	Standard	Experimental Group		Control Group	
		First	Second	First	Second
H	100	99	99	99	99
I	100	98	99	97	98
J	100	98	99	98	99
K	99	98	99	97	97
L	98	98	98	97	98
M	96	95	98	93	94
N	94	93	96	91	92
O	92	91	97	89	89
P	88	85	94	83	86
Q	84	79	86	76	80
Total		93	97	92	93

It took a school month to give the ten tests on alternate days. The next month the tests were repeated in the same order (increasing difficulty, H to Q) and with same time intervals except that the first five were given on succeeding days instead of on alternating days, thus using three weeks for the second tests.

Each spelling period on no-test days was devoted to regular work by the control-group classes, and to study of individual misspellings by experimental-group classes. Mimeographed cop-

ies of word and phrase lists were also furnished pupils of the latter group to use if they wanted to review the entire lists.

Campaign Results—Table I shows the percent of correct spellings in each list of words made by the experimental and control groups in each trial. Tables II and III give the same information for grades V and VI, respectively.

Table IV is a summary of results. Part (a) gives the data for the experimental groups and part (b) the data for the control groups. A comparison of the two parts of the table shows the control groups to have been somewhat poorer spellers on the initial tests and that their gain was quite negligible in comparison with that of the experimental groups.

Summary and Conclusions

In the experimental groups the increase in percent of correct spellings of 33,400 words by the fourth-grade pupils was 9 (84 to 93); of 48,050 words by fifth-grade pupils was 5 (90 to 95); and of 30,050 words by sixth-grade pupils was 4 (93 to 97). These percent increases of 9, 5, and 4 were made in one-fifth of a school year in regular spelling periods yet the mean percent increase in the Ayres scale on the same lists for an entire year's work between fourth and fifth grades is 7.5 percent and between fifth and sixth grades is 4.8 percent.

TABLE IV. SUMMARY OF RESULTS

(*a*) *Experimental Groups*

Grade	Number of Papers Each Trial	Number of Words Spelled Each Trial	Number of Words Misspelled Trial		Gain	
			First	Second	Number of Words	Percent
IV	668	33400	5527	2312	3215	58
V	961	48050	4552	2327	2225	49
VI	601	30050	12063	1045	1018	50
Total	2230	111500	12142	5684	6458	53
Percent Misspelled			11	5	29[a]	

(b) *Control Groups*

Grade	Number of Papers Each Trial	Number of Words Spelled Each Trial	Number of Words Misspelled Trial		Gain	
			First	Second	Number of Words	Per-cent
IV	652	32600	8824	7913	911	10
V	772	38600	5346	4908	438	8
VI	821	41050	3288	2821	467	14
Total	2245	112250	17458	15642	1816	10
Percent Misspelled			16	14	8*	

* Means number of words gained per pupil.

If pupils of the intermediate grades will eliminate 53 percent (Table IV) of their misspellings from the central half of the Ayres' list in one-fifth of a year's time simply because they know exactly where they stand, and that they will have a chance to prove the value of their efforts, are we not overlooking a great chance in the subjects, or divisions of a subject, which lend themselves to definite tests of development?

RECENT DEVELOPMENTS IN SILENT-READING TESTS

C. R. STONE
Principal, Gardenville School, St. Louis

Recent developments in the field of silent-reading tests have given us an equipment far superior to what we had two years ago. In particular, advancement has been made along three lines: (1) simplification in the testing procedure; (2) measurement of ability to read specific types of material; and (3) measurement of specific types of comprehension.

PART I. AN ADEQUATE EQUIPMENT OF SILENT-READING TESTS

Three types of silent-reading tests.—The most serviceable reading tests and scales now available for group testing may be classified into three types as follows:

1. Short time paragraph tests with units of equal difficulty.
   a. Monroe Silent Reading Revised
      Test I for grades III, IV, and V (three forms)
      Test II for grades VI, VII, and VIII (three forms)
   b. The Burgess Picture Supplement Scales 1, 2, 3, and 4 for grades III to VIII inclusive.
2. Long time paragraph tests with units of increasing difficulty.
   a. The Thorndike-McCall Reading Scale (ten forms) for grades III to XII.
   b. The Haggerty Reading Examinations
      Sigma 1 for the lower grades (one form)
      Sigma 3, Part III, for grades VI to XII (two forms)
3. Tests for measuring rate and comprehension in rapid narrative-reading.
   a. Fordyce's Scale for Measuring Achievement in Reading
      Test No. 1 for grades III, IV, and V
      Test No. 2 for grades VI, VII, and VIII
   b. The Stone Series of Narrative-Reading Tests
      Test for grades III and IV
      Test for grades V and VI
      Test for the Junior High School

Types 1 and 2 are alike in that the pupil may re-read each paragraph unit until he thinks he can indicate a correct com-

prehension response. In this respect, all these tests measure a study type of comprehension ability. Those of Type 3 measure the rate of continuous reading and the ability of the pupil to grasp the main points in a single reading. The kind of reading material used and the character of the comprehension measured are radically different from that used in the tests of Types 1 and 2.

The Monroe Tests and the Burgess Scales.—While the Monroe Silent Reading Tests (Revised) and the Burgess Scales are alike in certain fundamental characteristics, as indicated in the preceding outline, they are different in certain respects. The paragraph units in the Burgess Scales are all of the same type of material, while in any one of the Monroe tests there are narrative, descriptive, informational, and poetical units. The Burgess plan is in accord with the present tendency in attempting to measure ability to read a specific type of material with a particular scale.

However, the Burgess scoring plan limits materially the usefulness of the scales. The Monroe scoring plan shows rate in terms of words read per minute and comprehension in terms of the number of correct responses. This makes it possible to graph the results for a school and easily to determine the relative efficiency as to rate and comprehension. In the Burgess test, the pupil's score depends upon the number of correct responses, and the score in no way reveals the pupil's relative rapidity and accuracy. Consequently, it is impossible to graph the results for a school and thus to show the relative efficiency as to speed and accuracy.

The Burgess scores range from 0 to 100 according to the number of paragraphs read successfully. The fourth grade must read successfully seven paragraphs to obtain the standard median score of 50, while the fifth grade is required to read eight paragraphs successfully to secure the standard median score of 50. The requirement is similarly increased for each succeeding grade. This plan has several disadvantages in comparison with the simpler plan used by Monroe—*i. e.*, that of allowing a comprehension score of one for each unit read successfully. The

Burgess plan complicates the graphing of the results for a school or system. It makes more difficult the locating of a pupil's grade level of reading ability. It also makes difficult the establishment of age norms.

After considerable experience with the Monroe and the Burgess tests, the writer believes that the usefulness of both to the teacher and to the supervisor would be materially increased if the scoring plan would show the number of units read as the speed measure and the percent of these units which are read successfully as the comprehension measure. Under the Monroe plan of scoring, the pupil or the school that has a high rate will almost always have a high comprehension; whereas if the comprehension accuracy were shown as a percent, we should find many cases of high rate and low accuracy. What the supervisor needs to know about the school and the class and what the teacher needs to know about the class and the individual pupil is the relative efficiency in rate and accuracy and the scores should reveal this without the necessity of referring to the test papers or to tedious comparisons with standards.

The Thorndike-McCall Scale and the Haggerty Reading Examinations.—About the same time that Burgess had concluded that difficulty or power tests are not applicable in measuring reading ability McCall had decided that the difficulty or power test is the best measure of reading ability. The units in the Thorndike-McCall Scale are progressive in difficulty. The first paragraph is relatively easy, the next somewhat more difficult, and so on. The paragraph sections of the Haggerty tests follow the same plan. The two differ, however, in the types of response required. Haggerty follows the plan of requiring the pupil to select from a number of listed responses. This plan lends itself to economy of time and to a high degree of accuracy in checking and scoring. The Thorndike-McCall plan is to have the pupil write out the answers to questions. This plan has certain disadvantages. The checking of the answers is time-consuming and there is danger of inaccuracy on the part of the grader. Furthermore, this plan limits the test very largely to a type of question that tests the pupil's ability to get certain

definite detail, and consequently reduces to a minimum the use of questions that test the pupil's grasp of the major relational ideas or the essential total meaning of the paragraph.

The high cost of testing.—The Monroe and the Burgess tests are relatively inexpensive. The Thorndike-McCall tests are somewhat more expensive, and the Haggerty tests are very expensive. Testing material has been relatively more expensive than teaching material because testing material can be used but once while teaching material in the form of books and pamphlets can be used over and over. In the case of the Stone Series of Narrative-Reading Tests, while the initial cost is relatively high, the set of booklets and rate cards may be used over and over with incoming classes. Additional cost for answer sheets is very small.

Narrative-reading tests.—Besides measuring the ability of a pupil to do the study type of paragraph reading, the supervisor and the classroom teacher need to be able to measure his ability to grasp the main points in a narrative at a single reading and to be able to measure accurately the rate at which the whole narrative is read. The Stone Series of Narrative-Reading Tests[1] is designed for these purposes. At present there is available a test for the third and fourth grades, one for the fifth and sixth grades, and one for the junior high-school grades. Additional scales are in preparation. It is planned eventually to have one for each grade with two forms so that it will be possible to measure attainment and progress in any grade with a story of a difficulty level of that grade. This means that the teacher of each grade will have a testing equipment on narrative reading that will enable her to diagnose her incoming class and to measure progress after a certain period of instruction. This regular use of testing equipment the same as we regularly use the reading book, the writer conceives to be essential to the most effective type of teaching.

Part II will give a detailed account of the construction and the use of the Narrative Reading Test for the Junior High School.

[1]Published by The Public School Publishing Co., Bloomington, Ill.

Part II. The Stone Series of Narrative-Reading Tests

Difficulties involved in constructing a measure of narrative-reading ability.—Recent reading tests have been paragraph tests. It is much easier to measure ability to read paragraphs than ability to read a complete narrative. It is very easy to adopt a simple procedure in a paragraph test, while it is almost impossible to do so in a narrative test. In a speed test in paragraph reading, in such tests as the Monroe or the Burgess, the time taken to do the testing is very short, while it is almost impossible to give a narrative test in less time than it takes to give a power or difficulty test in paragraph reading.

The problem of obtaining an accurate measure of rate in some such terms as the number of words per minute has been a serious one in devising narrative tests. Gray invented a very accurate means of finding the pupil's rate of reading, but his plan involves individual testing; and this is too time-consuming for regular classroom purposes. On the other hand, classroom teachers and supervisors have become very skeptical of the commonly used plan of having the pupils read a minute or so and at a signal mark the last word read.

Another problem is so to relate the rate and comprehension tests that we may have a rate measure and a comprehension measure based on the same reading act. Only when this is done can we satisfactorily characterize the reader as slow and accurate, rapid and accurate, slow and inaccurate, or rapid and inaccurate. No group narrative test printed previous to 1922 solves this problem.

Construction of the Junior High School Test.—In the construction of the Junior High School Test a number of preliminary experiments were tried, an account of which it is impossible to give in this brief article. The following principles were kept in mind from the beginning:

a. The rate test should include for all pupils the reading of the whole narrative.
b. The test for the upper grades should be based upon a narrative of upper-grade level.

c. The comprehension test should measure the pupil's grasp of the leading points of the narrative.
d. Both the rate test and the comprehension test should be as simple in the testing procedure as possible.
e. The narrative should be of such length that no pupil would complete it in less than two minutes.

The first story tried was too easy; but a passage from Cooper's "The Spy" proved to be of eighth-grade level, and it was adopted. This narrative passage is about 1400 words in length. During the spring of 1921, eighth-grade pupils to the number of 563 were tested. The median rate was 234 words read per minute. On the comprehension test of ten questions, the median was 64, ten points being allowed for each question correctly answered. This extensive trial of the test followed preliminary trials and revisions. By this time the questions had been found to be approximately equal in difficulty. However, in constructing diagnostic charts for interpreting results it was found that ten questions did not give a sufficiently refined scale.

Very early in the work of constructing the test it was decided to have preparatory rate and comprehension tests precede the real tests. The preparatory tests are carried out in all details like the real tests. The unit of reading is, of course, much shorter.

The rate test is conducted as follows. The pupils are told that they are to read the narrative through from beginning to end, reading as rapidly as they can and at the same time get the story. Each pupil understands that he is not to look up until he has finished reading the story. Then he looks up and gets his rate from a card then being exhibited and copies it on his paper. A new rate card is exhibited each five seconds. The number on the card represents the number of words read per minute. A pupil, or some other assistant who has been trained for the purpose, exhibits the rate cards under the supervision of the examiner.

After the pupil has recorded his rate score, he proceeds immediately to answer the questions without being permitted to

re-read. Each question has five answers listed underneath it, only one of which is correct. The pupil answers the question by selecting one of the answers and indicating it on his answer paper. There are twenty questions in the tests as published. He is allowed as much time as he needs in answering the questions. The pupils are instructed to the effect that after each one finishes answering the questions, he is to take out a book and read until all have finished.

It takes about an hour to give the preparatory and the real tests. This, however, includes the checking and rechecking of the answers by the pupils—a procedure which makes it unnecessary for the teacher to spend time outside of school hours in scoring the papers.

Results of testing seventh- and eighth-grade pupils: mid-year, mid-grade.—Figure 1 shows the distribution of the rate scores of 320 eighth-grade pupils on a scale ranging from 500 to 100 words per minute. The graph shows that there were 15 pupils who read at a rate above 500. Only 1 pupil read at a rate as low as 100 words per minute. The mode is 220 and the median 245. We see that the curve is skewed somewhat to the lower end of the scale. Otherwise the distribution corresponds fairly closely to the normal curve of distribution.

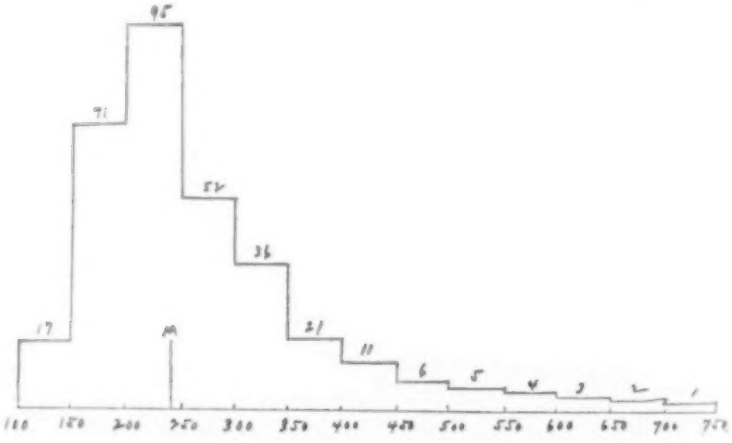

FIGURE 1. THE DISTRIBUTION OF RATE SCORES OF 320 EIGHTH-GRADE PUPILS

The usual wide range of attainment is clearly in evidence. The problem arises as to whether there is really such wide variability in the rate attainments of eighth-grade pupils or whether the marked variability is due to inherent defects in the measuring instrument. For instance, is the story of uniform interest to boys and girls alike, or do some of the pupils read rapidly because they are gripped by the story while others read slowly because the story does not appeal to them? It was the general opinion of those connected with the testing that the story is generally interesting to the pupils and about equally so to both boys and girls. However, a comparatively small number of pupils stated after the test that they did not care for the story.

It may be remarked in passing that in selecting the stories for the fifth- and sixth-grade test, preliminary experiments were carried out to aid in securing stories having a uniformly strong appeal to the pupils. Six stories were selected with a view of getting those that would be interesting to all pupils. All the pupils in a given room were asked to read six stories at one sitting and to put down afterwards the order of preference. The two stories having the highest degree of uniformity of appeal were selected for the test. The rate score and the comprehension score, in each case, is the average of the scores made on the two stories. This plan tends to decrease somewhat the wide variability in the pupils' scores.

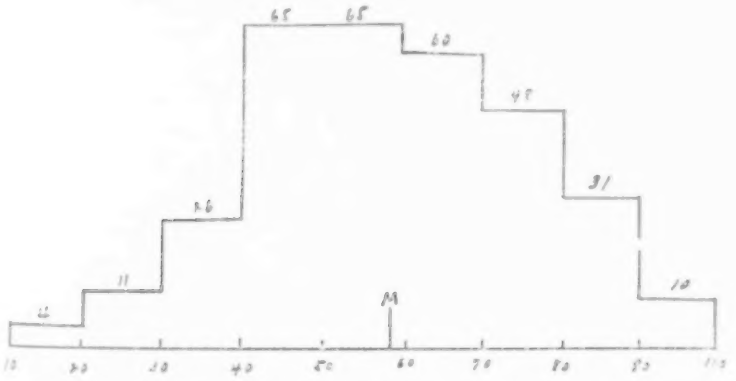

FIGURE 2. THE DISTRIBUTION OF THE COMPREHENSION SCORES OF 320 EIGHTH-GRADE PUPILS

Figure 2 shows the distribution of comprehension scores for the Junior High School Test on a scale ranging from 0 to 100, counting 5 for each of the 20 questions. The distribution curve follows rather closely the normal curve of distribution. It will be noted that no pupil made a perfect score, or in other words, that no pupil answered all 20 questions correctly. Likewise, no pupil made 0. Consequently all pupils were measured. The median is 57 or 11.4 questions answered correctly. Although it is considered that this story is of about eighth-grade level, the results for 274 seventh-grade pupils show that it is usable in the seventh grade also. The comparative results of the seventh and eighth grades indicate that the test will also be an accurate measuring instrument for the ninth grade.

In deriving the medians for Table I, data from only those schools which gave the test in both the mid-seventh grade and the mid-eighth grade were included. This was to insure more accurate rate and comprehension intervals between the seventh and the eighth grades.

TABLE I. MID-YEAR, MID-GRADE MEDIANS FOR ST. LOUIS

Grade	Rate	Comprehension	No. Pupils
VII	218	10.2	274
VIII	245	11.4	320

The use of the diagnostic charts.—To the classroom teacher, the most important fact that a test of this sort should reveal is the type of reading each pupil does. In almost any room or class of pupils, we find the slow accurate reader, the rapid accurate reader, the rapid inaccurate reader, the slow inaccurate reader, and all of the intermediate combinations of the two elements of rate and accuracy. The diagnostic chart is designed to show the teacher graphically just how her class distributes among nine rate and comprehension groups. The chart is constructed in such a way that on the basis of the 320 eighth-grade pupils tested approximately 50 percent will be in the middle comprehension division, about 25 percent in the inferior division, and about 25 percent in division labeled good. The same plan is followed with reference to the rate divisions.

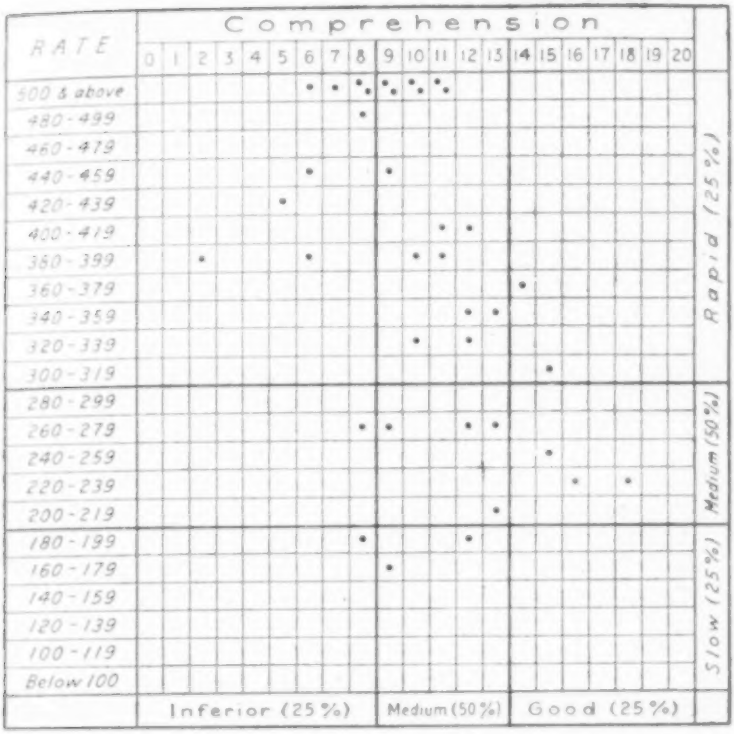

FIGURE 3. THE DIAGNOSTIC CHART FOR A MID-EIGHTH GRADE CLASS IN SCHOOL A. 37 PUPILS

Charts for three eighth-grade classes are reproduced representing three distinct types of classes with reference to reading efficiency. Figure 3 is the chart for a class in School A that is far above the average in rate and somewhat below the average in comprehension. The distribution indicates a small negative correlation between rate and comprehension. The fact is that this class in its training had been speeded up at the expense of comprehension. A number of these pupils should be told to read somewhat more slowly. This point should be particularly emphasized with the nine pupils falling in the rapid-inferior division. It will be noted that in this class the two pupils ranking highest in comprehension are medium-rate readers. The

medians in rate and comprehension for this class were 390 and 10.6 respectively. These may be compared with the tentative city medians of 245 and 11.4.

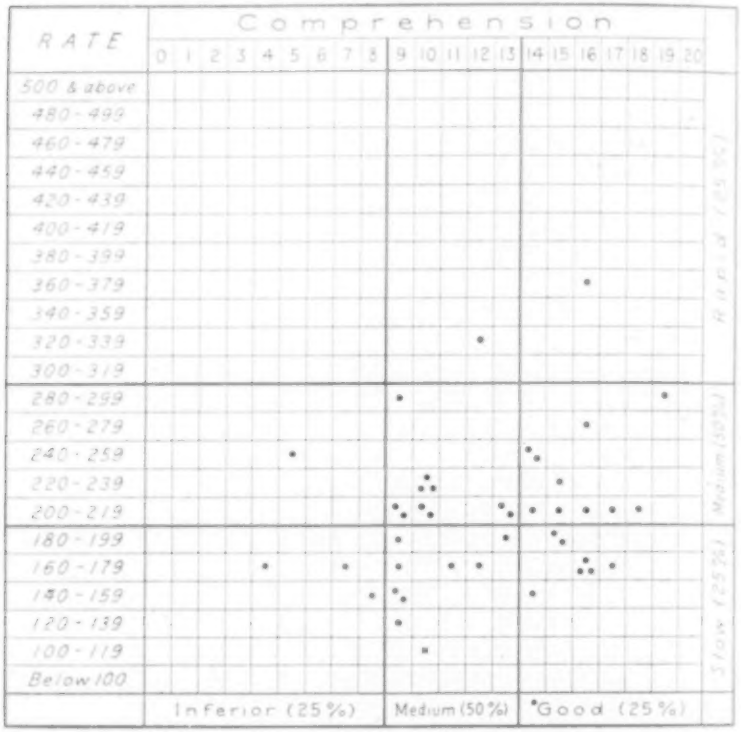

FIGURE 4. THE DIAGNOSTIC CHART FOR A MID-EIGHTH GRADE CLASS IN SCHOOL B. 42 PUPILS

The next chart, Figure 4, reveals a situation in School B that is radically different than that in School A in the same grade. This class is considerably above the standard in comprehension (median 13.0) and considerably below in rate (median 202). In the preceding chart the dots tended to distribute toward the upper left-hand corner, while in this chart they tend to fall toward the lower right-hand corner. There are seven pupils who are slow but good in comprehension. It is possible that this school has stressed thoroughness at the ex-

pense of speed. It appears that the training now needed by this class is of an opposite type from that required of School A.

It is not to be expected, of course, that each pupil or each class will perform according to some standard balance of speed and comprehension. The most efficient speed for one pupil may not be the best speed for another. The type of mind of the pupil will be an important factor. School B is in a community of relatively low social status. The intelligence of the pupils is below standard. The writer's experience leads him to believe that the pupil of low intelligence, in order to comprehend equally as well as the bright pupil, must read more slowly than the bright pupil. It is not meant by this that the slower one reads the better he comprehends. It is possible that those who are slow and inferior would comprehend better by reading faster. The problem is for the pupil to find the rate at which he comprehends best and then gradually to increase his rate as his ability to comprehend develops.

The class in School C, Figure 5, shows a greater variety in types of readers than either of the other two. The medians (2.34 and 10.0) are somewhat below the tentative norms. There are more pupils falling in the slow-inferior group than in either of the other two classes. Evidently these slow-inferior pupils are working at a great disadvantage if placed in the same class and given the same material as the pupils who are rapid or medium in rate and good in comprehension. The pupils of this class should be grouped according to their needs and should be given reading material of different levels as well as different types of training and experience.[2] Most of the pupils of this class should have training for developing speed.

Relation of rate and comprehension.—The teacher's manual for a set of readers planned for considerable amounts of silent reading contains the following statement: "The fastest readers usually make the best readers. Therefore, if you develop speed, you improve the quality of the child's reading." This reasoning not only does not accord with the results of research work in reading but it also fails to accord with the common experience

[2]For specific details see Stone, Clarence R., *Silent and oral reading*. Boston: Houghton Mifflin Company, 1922. 295 pp.

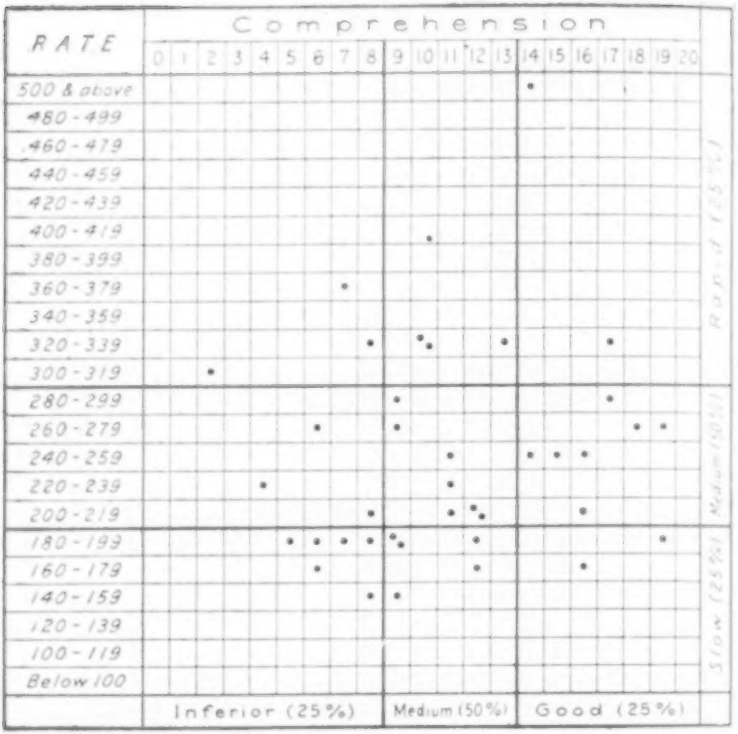

FIGURE 5. THE DIAGNOSTIC CHART FOR A MID-EIGHTH GRADE CLASS IN SCHOOL C. 38 PUPILS

of teachers. The data resulting from the application of the Junior High School Test show about what the St. Louis and the Cleveland Surveys showed, namely that the rapid-good pupils and the slow-poor pupils are about twice as numerous as the rapid-poor and the slow-good. However, this does not mean that increased speed brings improved quality. There is reason to believe that School A has developed speed at the expense of quality, while School B has not given sufficient attention to speed. No doubt very slow readers may comprehend better if they read somewhat faster. Probably all such pupils should be encouraged to speed up and should be given types of drill designed to produce a wider span of recognition per eye-pause.

On the other hand, many pupils have reached the point in the speed scale that for them gives the maximum efficiency in comprehension. For them to be speeded up without a corresponding development in ability to comprehend is likely to lower the quality of their reading. There is also the rapid-careless reader who should be given training that develops accuracy. A decrease in his speed will probably be advisable.

Pupils should be grouped and given reading instruction according to their needs. The diagnostic chart is designed to assist in determining their needs on the basis of each pupil's relative attainment in rate and comprehension.

Summary.—Each of the narrative reading tests has the following distinguishing characteristics:

1. There is a preparatory test, shorter than the real test, but carried out in every detail just like the real test.

2. There is a unique plan for measuring the rate involving the use of a set of rate cards. To get his rate score all the pupil is required to do is to look up when he comes to the end of the story and copy the number then being exhibited.

3. The comprehension test is upon the reading which has just been measured as to rate. It consists of twenty questions testing the pupil's grasp of the leading points of the story. The pupil selects one of the five listed answers for each question. The plan is simple and economical in time; it is an accurate measure of the comprehension of a narrative.

4. The tests are concerned with a specific type of reading material and they measure a specific type of comprehension.

5. Since the pupil enters his data on an inexpensive sheet, the test material may be used over and over again, thereby reducing to a minimum the per pupil cost of testing.

6. By using the diaguostic chart, the teacher can easily make practical use of the results in analyzing individual cases and in grouping her pupils for more effective instruction in accordance with individual needs.

7. The time required for testing a group of pupils including the checking of the answers and the recording of the scores is from 40 to 60 minutes. This means that the teacher is not required to do a lot of checking and scoring outside of school time.

SOME ESSENTIALS IN SCHOOL SUPERVISION[1]

E. J. Ashbaugh
Ohio State University

I suppose I may assume general agreement to the proposition that supervision is for the purpose of improving classroom instruction. If the supervisory force does not function through the teacher in increased efficiency in class work, its existence cannot be justified nor maintained. This increased efficiency must manifest itself through either a greater quantity of work of a given quality, an equal quantity of better quality or an equal quantity and quality at a lesser time expenditure. The supervisory group has tended to be recruited either through special training for this type of work or through selection of teachers who had had marked success in some particular line and who, it was therefore thought, might have ability to direct other teachers to acquire similar success. But whatever may have been the training or the method of selection, there are certain essentials which must be borne in mind if the work is to be satisfactory. It is the purpose of this paper to discuss a few of these.

The selection of the appropriate subjects by means of which to attain desirable educational objectives is a problem in itself. Tradition, current practice, and legal or institutional requirements are often more potent than the supervisory force which must deal with these subjects in the schools. However, an educational philosophy which interprets the content of the subjects in terms of present day needs and probable future needs may do much in determining the educational ends to be secured by means of these subjects.

Definite Goals

Assuming that these general subject fields are already determined, it seems to me that the first and most important essential

[1]An address presented at the Educational Research Department of the Ohio State Teachers Association, Cedar Point, Ohio, June 28, 1922.

in supervision is a rather accurate knowledge of definite goals of attainment. The best supervisors are not content today with intangible goals, ethereal objectives, indefinite generalities. Teaching reading in order that children may become cultured individuals is not a satisfactory statement for the supervisor, however true it may be as an ultimate objective. The supervisor must not only have an ultimate end in view but must have in mind a large number of intermediate stations definitely charted as guides in reaching this ultimate. I like to think of these intermediate goals as towns or cities along a well-marked automobile road. Each is definitely marked so we know when we have reached it. We know how far we have come since leaving the last place; we know what the next one is on the route and how far we must travel before arriving there.

Confining ourselves for a bit to the field already mentioned, to reading, let us seek farther for acceptable objectives. One says that we desire our children to learn to read in order to get the thought from the printed page. This is better than the former statement of aim, but it is still unacceptable from the standpoint of supervision. It is not a specific goal. Continuing the analogy to an automobile trip, the first statement (*i.e.*, of culture as an objective) would be like saying that the place to go is where the climate is pleasant, the people hospitable, and the roads good the year round. Such a place may indeed be desirable but the tourist who is directed in his journey only by such a statement will certainly be handicapped in proceeding on his journey. The second statement, namely as to getting thought from the printed page, might be likened in the same analogy to saying that the place to go is to the city in Missouri where there is satisfactory hotel and garage service at the least possible price. If that were the goal determined upon, it might serve to start the tourist off in the right general direction; but how would he know when he got there? The state of "thought getting" is nearly as large as the state of Missouri.

Further, in going to Missouri one might take any number of days largely at his pleasure; but in guiding the child to the desired goal in reading, one may not loiter, for the length of time

he may remain in school to profit by our guidance is problematical. It is certain that unless he reaches this goal in not more than four or five years, he will be greatly handicapped throughout the remainder of his career. Hence it is even more important that the goal be defined definitely and that landmarks be set along the way; for progress may then be noted and greater assurance may be had that the ultimate objective will be reached before it is too late.

A goal more specific than "somewhere in Missouri" would be Marysville, Mo. Similarly, a goal in reading more specific than "ability to get thought from the printed page" is "ability to read ordinary or non-technical material at the rate of not less than 240 words per minute and at the same time to get the thought so as to be able to reproduce, by answering questions or otherwise, not less than 95 percent of the ideas expressed in the selection."

This is a definite goal which may be set for attainment by at least half of the children by the end of the eighth grade. Having established such a goal, we may look for the intermediate goals or guides which will enable us to know what progress we are making toward this more remote objective. We may, for example, set up grade goals as follows:

Grade	Speed	Comprehension (percent)
I	50	50
II	95	60
III	125	70
IV	150	80
V	180	85
VI	200	90
VII	220	93
VIII	240	95

These goals, it is to be understood, are hypothetical and may be at considerable variance with those which should be established, although they are not greatly different from the norms of attainment which have been ascertained by workers in this field. Only tentatively, however, should such norms be considered goals,

only, indeed, until the necessary analytical and experimental work has been done to verify them or to establish new ones.

A vast amount of such analytical and experimental work is needed. In order to set a satisfactory goal for the end of the school course, such questions as the following need to be answered: At what rate and with what degree of comprehension will the average adult twenty-five or thirty years hence need to be able to read in order to live successfully his full measure of life as a worthy member of society at that time? We must set our goal in terms of the future. The school must prepare the child for adult life because it will be with those, who like himself are now enjoying greater educational advantages than their parents, that he must live and work. Doubtless this future need must be determined by estimate from the ascertained needs of the present. Even these, however, are thus far undetermined.

Other questions that will need to be answered are: What degree of overlearning must be accomplished in order that the desired residue may continue to exist after drill days are over? What adjustment of goals should be made because of variation in type of material? What other goals are legitimate objectives in reading besides speed and comprehension? What are definite goals in appreciation of literature, such as appropriateness of style, vividness of description, aptness of illustration, etc.

Or, to illustrate what I mean from another field, the supervision of the teaching of arithmetic cannot be done well with only general objectives. Here again the goals must be definite and a large number of intermediate points should be definitely charted. Teaching children to add well is vague until *how well* is indicated in terms of rate and accuracy in respect to a definite type of material. Instructing a teacher to develop ability to use common and decimal fractions without defining the ability more specifically leaves large leeway for the teacher either to be satisfied with too little ability or to insist upon the acquisition of ability far beyond any need which is likely to arise in the life of either child or adult.

What has been shown for reading and arithmetic is quite as true in the other subject-matter fields. Lack of definite goals,

clearly indicated by the supervisor, is readily recognized in the variation among the teachers of a system as to what constitutes satisfactory work on the part of the pupils.

Attaining the Objectives

The second essential follows logically from the first, namely, a rather definite knowledge of how to attain these objectives. Supervisors are, in theory at least, sources of help and inspiration for the teachers in the accomplishment of the tasks to be done. There has been much written and more said about the danger of over supervision, of the injustice in crushing out the teacher's initiative, of supervision making the teacher only a cog in the machine, etc. True supervision does none of these things, though it may prevent the teacher's initiative from straying far from the desired goals of her educational work.

True supervision certainly does not mean that the supervisor tells the teacher or demonstrates to her every new step along the way. In many of our schools the so-called penmanship supervisor does all the penmanship teaching that is done; the art supervisor does all the art teaching; the music supervisor all the music teaching. These are not supervisors; they are teachers assigned to a special subject in all rooms instead of to a group of subjects in one room. The true supervisor, operating directly upon the teacher and through the medium of the course of study, sets the definite goals in terms of quality and quantity of achievement. Page terminals in the textbook may indicate something of quantity, but they never define quality of work. Let us rule them out and substitute real worthwhile goals which define both quantity and quality.

Having defined the goal in the subject for the grade and perhaps indicated some of the landmarks along the way, the true supervisor will give the teacher an opportunity to exercise her initiative in attaining this goal. Detroit has been trying out a scheme whereby the supervisor is prohibited from visiting the classroom or making any suggestion to the teacher except upon request. This not only gives the teacher an opportunity to exer-

cise her initiative but it also places responsibility squarely upon her either to attain the goal alone or to request help. If she succeeds alone, the glory is hers and the supervisor has been left free to work where needed. If she realizes she needs help to attain this goal and asks for it, she is in the best condition to profit by the help the supervisor can give and hence the time of the supervisor is economized. If she does not call for help and fails to attain the objective set, the responsibility rests squarely upon her and indicates that either she can not judge accurately the progress of her pupils or over-estimates her own ability or refuses to acknowledge her need. In either of these cases she certainly needs supervision.

Regardless of whether the responsibility for recognition of needs be placed upon teacher or supervisor, whenever the need exists, the supervisor should be able to furnish it. Just how this aid should be given may vary in terms of the personality of supervisor, teacher, type of need, etc. The supervisor may demonstrate, taking the class before the teacher and illustrating the working procedure by which the desired results may be obtained economically and efficiently. She may, in personal private conference with the teacher, tell how the teaching process should be carried on or give suggestions which will supplement the ideas which the teacher has already been utilizing. She may, either in conference or by letter, direct the teacher's attention to available literature in which the teacher will find specific help in her difficulty. The true supervisor may use any or all of these means to secure the accomplishment of the goal but in all cases she must know fairly definitely how to attain the goal or where the information may be found.

Checking Attainment

The third essential is closely associated with the first two. Efficient supervision is dependent upon the supervisor's knowledge and use of valid means for checking the extent to which the goals set forth have been attained. To set forth ends to be secured without ascertaining whether or not they have been

attained is folly. It is worse than folly, since it will quickly break down the effort on the part of the teachers to attain these objectives. The teacher who questions the validity of the objective or who has ends of her own which she considers more important will quickly cease to strive for the former when she learns that no check is made of her attainment of the objective in question.

Probably the best reason for the use of adequate means of checking attainment is exemplified in the old adage, "Nothing succeeds like success." Children and teacher alike wish to know of the success of their efforts. If the objective for eighth-grade reading has been set at 240 words per minute with 95 percent comprehension, every eighth-grade child and teacher will be interested in knowing whether that goal has been reached or exceeded. Every child who has passed it will feel the glow of successful effort. This will be shared by his teacher. The placing of goals upon an objective basis removes the personal prejudice charge on the part of those who fail. Supervisors may well expend a part of their time and ingenuity in devising objective tests for measurement of success in attaining their definite goals, and they must certainly utilize those tests which are already available.

But it is more important that progress be known before the end of the year than at that time. Diagnostic tests which not only show status attained but also reveal weaknesses are of supreme value in supervisory work. To the extent that difficulties are discovered early, and to the extent that no time is lost in futile effort and no wrong habits are formed, progress will approach maximum rate. The teacher should be keenly alive to these facts and ready to use the means available to check on these items. But it is the function of supervision as it is organized in most systems to see that these items are cared for whether the teacher is alive to their importance or not. Hence the supervisor must be aware of and use these instruments for checking progress.

Supervision without definite goals and objective means of checking progress is likely to be of little value. A state wide

study of handwriting in Iowa a few years ago shed a little light upon this phase of the topic under discussion. Rural schools in that state have practically no supervision, since one county superintendent for 150 to 225 one-room schools scattered over an area of 500 to 900 square miles can not possibly supervise much, even if he had nothing else to do. And he certainly has many other duties. On the other hand, all cities of 10,000 or more had, and had had for a number of years, special supervisors of this subject. A single system of penmanship instruction was in use in most of these cities. But definite goals of attainment in terms of specific rates of speed and qualities of form were not in use and objective means of checking attainment were also lacking. When, therefore, the samples of writing were carefully rated by the Ayres scale, the results revealed that the median quality of handwriting in all grades in city schools averaged less than one-half step on the scale better than the quality of writing of the rural children and in the third and fifth grades the quality of writing of rural children was actually better than that of city children. A similar survey in the fundamentals of arithmetic showed an advantage for city children of only about one-half the average inter-grade advance. Such lack of significant difference between the achievement of supervised and non-supervised schools points clearly to ineffective supervision.

Opportunities for Research

These three essentials call for much research work and supervisors in daily contact with both teachers and children are in a particularly strategic position to conduct at least certain phases of it. The definite goals by subject and grade can best be determined in the schoolroom. Supervisors may well set the goal high and then utilize all available knowledge in method and psychology of learning to set new records of accomplishment. The little that has been done has shown conclusively that the progress that is normally made in rate of reading in two years, can be attained in less than half that time with no loss of degree of comprehension. Richardson's experiments in spelling and arithmetic have shown that "campaign" methods enable one to

speed up the process of attaining specific goals very materially. Analysis of social needs such as the work of Horn in spelling and the social sciences and of Wilson in arithmetic are fundamental to the determination of the proper content of our curriculum. Such work as that of Charters in language errors lays the foundation for specific application of corrective measures. Every supervisor could add to the knowledge now available if she would secure the cooperation of her teachers and their pupils in a small but definite bit of research work each year.

This naturally suggests another essential in supervision, namely the ability of the supervisor to secure the whole-hearted cooperation of her teachers. This is secured only through the demonstration of a sympathetic, helpful attitude. Whenever a teacher discovers that the supervisor is a sympathetic source of help and inspiration, her cooperation is at once guaranteed. The goal set up by the supervisor may be new and apparently unattainable, but she will give her utmost effort knowing that the supervisor will gladly lend a helping hand and will either show her how to attain the proposed objective, or cite her to the literature which will explain. The devices suggested may be strange but they will receive consideration because she will know that the supervisor is as vitally interested in her success as in the success of these boys and girls. She will also know that a meed of praise will be hers when the checking is done and when it is seen that aspiring to the goals and following both her own best initiative and the advice of her supervisor the desired results have been attained.

The program suggested here is applicable in any place where there is any supervision at all. If you are a superintendent in a small place where all supervision must be done by you, it is still possible for you to follow it. If you feel you can do more, select one subject; make yourself acquainted with the literature in the field so that you may set up definite goals of attainment which will be in line with the best that is known on the subject, be able to cite your teacher who needs help to books and articles which will aid her in attaining these goals; use the means available to check results; and secure the sympathetic cooperation of

your teachers in doing a bit of real research on this problem in your own schools. It means some time and effort and doubtless you feel that your time and effort are about all taken now. But you owe it to your school, to your teachers, and to yourself to do this. It means service of the best type to both school and teachers, and it means professional development of yourself at the same time.

CAN TEACHERS SELECT BRIGHT AND DULL PUPILS?

G. F. VARNER
College of Education, University of Minnesota

Scientific investigation of the use of rating scales has shown that ratings are not reliable measures of human traits. Rugg,[1] in discussing the rating of officers in the United Staes Army, states that: "It was very improbable that an officer was located within even his proper 'fifth' of the entire scale by an 'official' rating." While these findings discredit teachers' ratings of pupils' abilities it might be possible that teachers can select those pupils who represent the two extremes of a trait. This paper is a discussion of an attempt to measure the ability of teachers to select the extremes of general intelligence. Intelligence was chosen in preference to other traits because: (1) It is the only trait for which satisfactory objective measures are available with which to check the ratings. (2) Teachers' estimates are often used in the selection of bright and dull pupils to be put into special classes. (3) If we believe that the bright children should be given a chance to progress more rapidly than the average children, then it becomes important to know whether the teachers can, without the use of intelligence tests, select those who might profit by such an opportunity. (4) Unless teachers can select the bright pupils, they will not know how to advise pupils intelligently in the selection of a vocation. Unless a teacher is able to estimate which are the bright children, some children will be advised to enter professional occupations when their ability is not great enough to warrant their preparation for such a life work.

From the data presented here an attempt will be made to answer the following questions: (1) How reliable are teachers' selections of bright and dull pupils? (2) Is it more difficult to select the bright pupils than it is to select the dull pupils?

[1] Rugg, Harold O., "Is the rating of human character practicable?" *Journal of Educational Psychology*, 12:436, November, 1921.

(3) Does it become more or less difficult to select bright and dull pupils as they become older?

In the schools of St. Paul, Minnesota, the teachers select the upper and the lower twenty percents of their pupils in general intelligence in order to supplement the intelligence test scores in determining which pupils will be put in special classes. The material used in this investigation was obtained from the St. Paul schools in 1920. It concerns six eighth-grade classes containing 206 pupils and three second-grade classes containing 80 pupils. The teachers were given a definition of intelligence and were directed to select the best and the poorest twenty percents in their grades. The eighth-grade pupils were tested with the Haggerty Intelligence Examination, Delta 2 and the second grades were tested both with the Haggerty Examination, Delta 1 and with the Pressey Primer Scale. One of the second grades was not tested with the Delta 1, so in this case the results were for the Pressey test only.

The scores of all the pupils were translated into intelligence quotients and the teachers' selections were compared with the intelligence quotients, as well as with the actual test scores. It was found that the teachers' selections compared more favorably with the intelligence quotients than with the actual scores. So for this report the intelligence quotients are taken as affording the mental classification to use and the results are reported for the intelligence quotients only. Since the teachers were directed to select the highest and the lowest pupils in their grades regardless of the general intelligence level of their pupils, the comparisons must be made for each grade separately. The comparisons were made by dividing all of the pupils of each grade into twenty percent classes according to their intelligence quotients. The bright and dull pupils as selected by the teachers could then be located within the twenty percent class to which they belonged according to the test classification. The results of this comparison are shown in Table I for the eighth-grade classes and in Table II for the second-grade classes. They are summarized by percents in Tables III and IV.

TABLE I. A COMPARISON OF THE UPPER AND LOWER TWENTY PERCENTS OF PUPILS AS SELECTED BY THE TEACHERS WITH THE INTELLIGENCE QUOTIENTS AS FOUND FROM THE HAGGERTY DELTA 2 INTELLIGENCE EXAMINATION. (EIGHTH GRADE.)

School	Selected by Teachers	Classification by Intelligence Quotients				
		First 20 percent highest	Second 20 percent	Third 20 percent	Fourth 20 percent	Fifth 20 percent lowest
I.	Upper 20 percent.	4	2		1	1
	Lower 20 percent.			1	1	5
II.	Upper 20 percent.	4	2	2		
	Lower 20 percent.				3	3
III.	Upper 20 percent.	1	1	2	3	
	Lower 20 percent.				2	5
IV.	Upper 20 percent.	3	2	3		
	Lower 20 percent.		1		1	3
V.	Upper 20 percent.	4	2	1		
	Lower 20 percent.				1	3
VI.	Upper 20 percent.	4	3	1	1	
	Lower 20 percent.			2	2	5
Totals	Upper 20 percent.	20	12	9	5	1
	Lower 20 percent.		1	3	10	21
Per-cents	Upper 20 percent.	42.55	25.53	19.15	10.63	2.13
	Lower 20 percent.		2.63	7.90	26.32	63.17

TABLE II. A COMPARISON OF THE UPPER AND LOWER TWENTY PERCENTS OF PUPILS AS SELECTED BY THE TEACHERS WITH THE INTELLIGENCE QUOTIENTS AS FOUND FROM THE HAGGERTY DELTA I AND PRESSEY INTELLIGENCE EXAMINATIONS. (SECOND GRADE.)

School	Selected by Teachers	Classification by Intelligence Quotients				
		First 20 percent highest	Second 20 percent	Third 20 percent	Fourth 20 percent	Fifth 20 percent lowest
I	Upper 20 percent.	1	2	2	4	1
	Lower 20 percent.		2	2	1	4
II	Upper 20 percent.	2	1	1	1	
	Lower 20 percent.			1		4
III	Upper 20 percent.	1	1		1	
	Lower 20 percent.		1	1		1
Totals	Upper 20 percent.	4	4	3	6	1
	Lower 20 percent.		3	4	1	9
Percents	Upper 20 percent.	22 22	22 22	16 65	33 33	5 55
	Lower 20 percent.		17 24	23 53	5 83	52 94

Table I shows, for example, that of the eight pupils selected by the teacher of School I to be in the upper twenty percent class in intelligence, four were put in that class by the intelligence quotient classification, two were put in the second class, one was put in the fourth class and one was put in the fifth or lowest class. Of the seven pupils selected by the teacher to be in the lowest twenty percent class in intelligence, five were classified in the same class by the intelligence quotients, one in the class above and one in the middle class. Table II gives the same data for the second-grade classes.

TABLE III. A COMPARISON OF THE ABILITY OF TEACHERS TO SELECT BRIGHT AND DULL PUPILS. (EIGHTH GRADE.)

	Percent of Pupils in Same or Different Classes According to Intelligence Tests				
	Same class	One class removed	Two classes removed	Three classes removed	Four classes removed
Upper twenty percent according to teachers	42.6	25.5	19.2	10.6	2.1
Lower twenty percent according to teachers	63.2	26.3	7.9	2.6	0.0

TABLE IV. A COMPARISON OF THE ABILITY OF TEACHERS TO SELECT BRIGHT AND DULL PUPILS. (SECOND GRADE.)

	Percent of Pupils in Same or Different Classes According to Intelligence Tests				
	Same class	One class removed	Two classes removed	Three classes removed	Four classes removed
Upper twenty percent according to teachers	22.2	22.2	16.7	33.3	5.6
Lower twenty percent according to teachers	52.9	5.9	23.5	17.2	0.0

Tables III and IV are summaries of Tables I and II. They show the percent of pupils concerning whom the teachers' estimates agreed with the test scores. They also show the percent of pupils concerning whom the teachers disagreed with the test results and the amount of the disagreement. Table III shows that of those selected by the teachers to be in the upper 20 percent of all the eighth-grade classes, about 43 percent were likewise placed in the upper 20 percent of these classes according to their intelligence quotients. If we may assume that the test results are correct, 25.5 percent of the pupils were misrated by the teachers to the extent of one class, 19.2 percent to the extent of two classes, etc. Table IV gives the same summary for the second-grade classes.

We may now turn to the three questions which we raised near the beginning of our article. From Tables III and IV we observe that teachers agreed with test results in from 20 to 60 percent of the cases on which they passed judgment. Their disagreements, however, were much closer for dull children than they were for bright children. In the second grade about 53 percent and in the eighth grade about 63 percent of the teachers' estimates that pupils belonged in the lowest 20 percent were verified by the intelligence quotients. On the other hand, when the teachers judged that pupils belonged to the best 20 percent of their classes, they agreed with the intelligence quotients much less frequently—in but 22 percent of the cases in the second grade and in but 43 percent in the eighth grade.

It appears to be appreciably more difficult for teachers to judge the mentality of pupils in the second grade than in the eighth. Table IV when taken in comparison with Table III exhibits this fact so far as the teachers who participated in this investigation are concerned. In round numbers 20 percent more of the bright and 10 percent more of the dull children were correctly selected by the teachers in the eighth grades than in the second grades. In another investigation by the writer it was found that the teachers' estimates of the intelligence of 171 third-grade pupils correlated 0.43 with intelligence quotients while the estimates of intelligence of 89 seventh-grade pupils correlated 0.67 with the same measure. These correlations substantiate the conclusion to be drawn from this investigation, namely, that it becomes less difficult to estimate the intelligence of pupils as they become older.

To the extent that the results of intelligence tests may be accepted as valid and the judgments of these teachers may be regarded as typical, the following statements may be offered:

Summary

1. Teachers can select 20 to 40 percent of the bright pupils in their grade and from 50 to 60 percent of the dull.

2. It is more difficult to select the bright pupils than it is to select the dull.

3. It becomes less difficult to select the bright and the dull pupils as they become older.

4. Selection of pupils for special classes by teachers' estimates is a very unreliable and unsatisfactory method of procedure and should not be used if it is possible to use an intelligence test. In the selection of pupils to enter classes for gifted children, teachers' estimates of intelligence are of practically no value.

5. From this study of teachers' ability to select the intelligent and the unintelligent children, inferences may be drawn as to their ability to select pupils according to other character traits.

THE MENTAL AGE OF ADULTS

EDWARD A. LINCOLN
Harvard University

The results of mental testing in the United States army during the late war left one vexing problem for the psychologist. The data, taken at their face value, indicate that the mental age of adults is not sixteen, as assumed by the Stanford Revision of the Binet test, but only about thirteen and a half. This point has become somewhat controversial, and has been recently attacked by Doctor Terman, who states, in the Journal of Educational Psychology (vol. XII, no. 6, p. 330), several reasons which, he believes, invalidate the inferences which may be drawn from the Army results. It is the purpose of this paper to inquire into the validity of Doctor Terman's objections to the inferences from the army results, and also to seek further light on the whole problem.

Among the objections which Doctor Terman offers, one of the most potent, if it can be shown to hold, is that "the 653 enlisted men[1] for whom the average mental age of 13.4 was found cannot be taken as representative of the entire draft army." The assumption is that these men were of inferior intelligence as compared with the draft army as a whole since, "A disproportionate number of them were from southern and semi-southern camps where, according to all the results of Army testing, average intelligence is lower than in northern and western states."

This assumption is not supported by the facts, which are shown in Table I. Here we find that the 653 group, when com-

[1]In order to obtain data regarding the comparative scores of the several examinations used in the Army, requests were sent out from the Surgeon General to the psychological examiners in 12 camps that 100 in each camp, relatively unselected, be given Alpha, Beta, and the Stanford-Binet. These examinations were given for the most part in August, 1918, and of 748 men examined, 653 were born in English speaking countries. For full report see Yerkes, R. M., *ed.*, *Psychological examining in the United States Army.* (Memoirs of the National Academy of Sciences, V. 15, 1921) p. 384 ff. Cited as "Army report" in this article.

pared with the white draft as a whole, has a distinctly higher percent of A and B men, regardless of the examination upon which they are graded. The Alpha and Beta examinations also give the 653 group a larger portion of C+ men. If the sums are taken of the percents of men making grades above the average, that is, all A, B, and C+ men, the comparison again favors the 653 group. Table I also shows that in regard to the grades below the average there is little difference.

TABLE I. PERCENTS OF WHITE DRAFT[a] AND OF 653 GROUP MAKING EACH GRADE

	Grades								
	D−, E	D	C−	C	C+	B	A	Above C	Below C
White Draft (94,004 cases)	7.1	17.0	23.8	25.0	15.2	8.0	4.1	27.3	47.9
653 Group									
Alpha	18.1	8.4	19.3	21.3	18.2	8.9	5.8	32.9	45.8
Beta	6.3	21.0	21.8	20.8	15.9	9.2	5.1	30.2	49.1
Stanford-Binet	8.0	15.6	23.0	22.4	13.2	11.9	6.0	31.1	46.6

[a]Figures for white draft from several tables in Army report, for example, Table 419 on page 865. Those for the 653 group are from Table 57, page 393.

Comparisons may also be made on the basis of the combined scale scores for the two groups, which may be found on Tables 332 and 333 on pages 789 and 790 of the report. In spite of the fact that the 653 group has no scores on the combined scale above 21.9 while the White Draft has 4.67 percent of its scores above this point, the smaller group has a slightly higher median; 13.5 as compared with 13.46 for the White Draft. This difference persists when translation is made into mental age; the median for the 653 is 13.25 years, as against 13.15 for the group from the principal sampling. It may be fairly argued, of course, that these differences are so small that they are not significant. However, this evidence, though slight, is cumulative, and with the facts in the previous paragraph it seems to indicate unequivocally that we cannot assume that the 653 group was inferior in general intelligence to the draft army as a whole.

A further objection which has been raised to the use of the Army results is the assumption that "the draft army itself was

certainly not representative of the male population between the ages of 21 and 31," largely because it is believed that deferred classification and exemptions were, for the most part, granted to the men of superior intelligence. Data are not available for the proof or disproof of this assumption, but it is possible to point out some rather fundamental considerations which Doctor Terman omitted to take into account when he made it.

The Army report, on page 830, presents figures concerning the exemptions and deferred classifications which are gathered for convenience into Table II.

TABLE II. PERCENTS OF EXEMPTION AND DEFERRED CLASSIFICATION FOR VARIOUS REASONS

Dependency	51.1
Alien allegiance	13.16
In service	8.9
Disability	7.75[a]
Occupation	6.8
Religion	0.5

[a]The report says these men were "totally and permanently physically, mentally, or morally unfit..." This figure does not agree with the report of the Statistics Branch of the General Staff which states that from 15 to 50 percent were rejected for physical reasons. (p. 20-21).

Let us analyze as well as we may each of these groups, and find, if we can, any reasons for believing them to be composed of individuals who vary to any great extent on an intelligence basis from the population as a whole.

Over one-half the exemptions and deferred classifications were for dependents. Probably most of them were granted to men who were married and were supporting their wives and children. Is this group superior in intelligence to the population as a whole? It scarcely seems that we can make this assumption, and it would be hard indeed to prove it. Such evidence as there is seems to point the other way. The more intelligent man stays longer at his education, and is thus kept from early marriage. Moreover, he usually wishes a permanent place on a relatively high level before he takes the responsibilities of family life. The intelligent man is also more often successful in his work, and many of the men of this class were in such a position

that they could not consistently claim exemption on the ground that they needed to continue their work in civil life in order to support their dependents. On the other hand, the less intelligent man, without ambition or foresight, marries early and has a large family. Thus it seems probable that the exemptions on this ground did not keep the intelligent man out of the draft army, but rather pulled him in, since it was more likely that he was unmarried, or if married, that his family was independent of his continued occupation in civil life.

Exemptions in the second largest group were for alien allegiance, that is, for those men who were not citizens of the United States. Are these men above the average intellectually? The Army report (part III, chapter 6, page 693) gives the results of the examination of some 12,000 foreign-born men who were included in the draft army because they had either become citizens or waived exemption. They were, as a whole, distinctly inferior to the rest of the draft army, as the figures in Table III will show. It is, of course, impossible to be sure that the foreign-born men who were drafted were on the same intelligence level as those who were not, but there seems no reason to believe that great differences existed. The only bit of evidence

TABLE III. PERCENTAGE DISTRIBUTION OF LETTER GRADES MADE BY FOREIGN-BORN MEN IN THE DRAFT ARMY

	E, D−	D	C−	C	C+	B	A	Number of Cases
Foreign	14.8	30.8	16.5	26.6	7.3	2.9	1.1	12,407
Whole Draft	7.1	17.0	25.0	23.8	15.2	8.0	4.1	94,004

of which the writer is aware is that the conscientious objectors who refused to serve in the army because they claimed not to be citizens, because they were enemy aliens, or because they had enemy alien relatives fighting on the other side, were, as a group, of very low intelligence.[2] This would seem to indicate that the more intelligent foreigners came into the army, especially when

[2]Army report, p. 802, section 2.

considered with the fact that the men who were naturalized, and therefore liable to the draft, were in all probability superior to the average of the foreign-born. The same is likely to be true of those men who waived exemption. Thus in this field the weight of evidence seem to be that if the general population was not well represented by the draft army it was because the men in the army were of a higher, not a lower, intelligence level than the men who were not drafted.

Doctor Terman believes that the men exempted from the draft because they were already in the service formed a superior group because, "these volunteers included a disproportionate number of college graduates, college students, recent graduates of high schools, high-school students, and high-minded youths from the better classes generally." This may be true, but it cannot be accepted unless accompanied by adequate proof, first, that there *was* a disproportionate number of college, high-school, and high-minded youths in the group, and second, that those who volunteered were superior to the men of like education and high-mindedness who waited to be drafted. The declaration of war found some 200,000 men in the United States Army, approximately two-thirds of them regulars and the rest National Guardsmen, who had been called for service on the Mexican border.[3] Surely no one familiar with the type of man who joins these organizations will contend that, with all their virtues, these men constitute any part of "the intellectual cream of the country."[4] This group comprises about a third of those exempted because of prior enlistment. The others were men who had enlisted when the call for volunteers was issued, or sometime before the enlistments were closed. Can we declare that the men in this group who came, say, from Harvard University were more

[3]Report of statistics branch of the general staff, p. 16.

[4]In the preliminary tryout of examination *a* when the scores ranged from minus 188 to plus 226 the following results were obtained at Syracuse, Regulars —73; at Indianapolis, Regulars and National Guard —47; at Nashville, National Guard —35. These scores correspond to 115, 121, and 153 on the later examination *a* of 414 points, on which the median score of the enlisted men in the 4 trial camps was 151. Mosquito fleet men at Brooklyn tested +88 (176), but their inclusion in one group with the army men leaves the score —26 (162). Army report p. 315, Table 1, and p. 491, Table 145.

intelligent than the Harvard students and graduates who were later drafted? Decidedly not without some pretty clear evidence. There may be reason, indeed, to suppose that if any selective factor was acting it worked the other way, and that unsuccessful men, the adventurers, and the like, were the ones to rush into the army, while the superior men waited to be called. It is not without import, surely, that the percent of desertions from the army was smaller among the drafted men than among those who volunteered.[5]

The exemptions for industrial and occupational reasons form a rather small group, but it is worth while to consider them as we have done the others. From the positions they held it seems quite legitimate to agree with Doctor Terman that they were superior individuals, but this argument does not take account of the fact that a man's position in his chosen field does not depend upon his mental ability alone. It may depend largely upon his father's ability, or upon the social, political, or financial connections of himself, or of any member of his family. There were charges during the operation of the selective service act that illegitimate exemptions were secured in this way. If we leave these considerations out, however, it is certainly clear that the attainment of the positions of leadership is due in a large measure, at least, to the length of time that a man has been working in his particular line. Men do not become "managers," "directors," "comptrollers" and the like until they have had a considerable degree of experience. Probably, then, the men who were exempted in this group were not more intelligent than their fellows, they were simply older.

Some evidence for this conclusion appears in the Army report in the chapter dealing with the intelligence ratings of occupational groups.[6] It is here stated that, "consistent significant differences in mental ability between apprentices, journeymen, and experts are not present." In only a few of the occupations do the experts score higher than the journeymen, and the journeymen higher than the apprentices, and many of the dif-

[5] Report of statistics branch of the general staff, p. 17.

[6] Army report, p. 831 ff.

ferences which do exist are small. Frequently the apprentices scored higher than the experts. This is exactly what would be expected from the considerations in the preceding paragraph. The exempted men were no more intelligent than the beginners; they were simply farther along in age and experience.

The men exempted for physical, mental, and moral unfitness surely cannot be considered a part of the "intellectual cream of the country." They might possibly be considered as average, but even this is extremely doubtful. In the first place, this group contained those who were rejected on the basis of inferior mentality, and this fact alone would probably bring the group below the average. We know, too, the tendency of the mentally inferior to be physically inferior also, and undoubtedly some of the men rejected for physical reasons were of inferior mental calibre as well. For some groups of physical rejections we may be sure that the average mentality was low, since the Army report points out the mental inferiority of venereals and men infected with hookworm.[7] The sum of the evidence seems to show that exemptions in this group left the inferior men in the general population and put the superior men into the army.

The objection that the low scores were caused by unfavorable conditions of examination is refuted, in part at least, by the finding that recently inoculated men did as well as men whose inoculation was taken at a time when it was impossible for it to function at the time of the examination.[8]

Further arguments by Doctor Terman deal with the form of the Stanford-Binet which was given in the army, and the training of the examiners. These he presents without evidence. The abbreviated scale, he believes, gives lower mental ages for adults than the complete scale, and untrained examiners are likely to be severe in their scoring. On the first point it should be considered that the omitted tests were for the most part those which are linguistic in their nature. They were omitted because it was supposed that many of the men who would be tested by

[7] *Ibid.*, pp. 810-11.

[8] *Ibid.*, p. 809.

the individual methods could not do themselves justice on tests of this nature. Further, the tests remaining were, in many cases, those in which the men were able to make better records because of experience rather than mentality. The making-change test is a good example in this connection. It is a nine-year test in the Stanford revision, but it was passed by a large percent of the adults who tested only eight.[9] There is at least a possibility, then, that adults test higher rather than lower on the abbreviated scale, and it is almost certain that such would have been the case with most of the men in the army who could not be marked fairly on the basis of Alpha.[10]

In connection with the second point it must be remembered that when the short army examination was made there were omitted many of the tests in which the examiner has to use his own judgment in scoring the response. For example, the omitted items included all the vocabulary tests, the meaning of abstract words, and the differences between abstract words. The fact of general severity in marking by examiners of small experience must also be called into question. It has never been noted in the training of students at the Psycho-Educational Clinic of the Harvard Graduate School of Education. We have found severity and leniency in marking to be largely a matter of the individual examiner.

The most serious objection to Doctor Terman's arguments, however, arises not so much from these arguments themselves, but rather from the fact that he failed entirely to take account of the striking evidence which arises from another source. He says nothing whatever concerning the fact that in many cases where the Alpha examination has been given to school children the thirteen-year-olds have equalled or surpassed the performance of the army men. The results of these testings in schools are especially significant because they are open to none of the objections which are raised in connection with the Stanford re-

[9] *Ibid.*, p. 408. Also, Lincoln, E. A., and Cowdery, K. M., "An abbreviated mental age scale for adults," *School and Society*, 10: 708, December 13, 1919.

[10] The Army report on p. 399 discusses the reliability of the four-test scale, and seems to justify its use.

vision, and we need not even take account of the one unquestionably valid point raised by Doctor Terman, namely, that the revision may be too difficult in the upper ranges.

This superiority of thirteen-year-olds over army men appeared when the Examination *a*, the original group examination used in the army, was given to groups of California school children. It is to be remembered that the scores of all children foreign-born or of foreign-born parentage were eliminated from the distribution, but to offset this, in some measure at least, is the fact that the scores for the enlisted men are for only those who were considered literate. Care must also be taken not to fall into the error of considering the enlisted men as a group which scored low partly because it included no officers, since many of the men in this group were afterwards sent to officers' training schools and commissioned.[11]

In the Army report the following results of Examination *a* are recorded:

Median score	399	12-year-old children, grades 3–9	155[a]
"	389	13-year-old " " "	182[a]
"	274	14-year-old " " "	207[a]
"	56,140	enlisted men, literate, from Camps Devens, Dix, Lee, and Taylor	161[b]

[a] Army report, p. 538, Table 186.
[b] Army report, p. 491, Table 145.

These figures seem to indicate that, even after considerable allowance for all sorts of factors, the army median cannot be considered to equal or exceed that of the thirteen-year-old children.

There are not in the Army report so far as the writer is aware, any results of the scores made on Alpha by school children. This examination, however, has been used in a number of schools since its publication, and the results have given further evidence that the mental age of adults is on a lower level than we had supposed before the army tests gave us a hint on the situation. The median Alpha score for 67,254 literate white men

[11] Doctor Terman falls into this error when he states that all officers were excluded from the 653 group.

is 60.5,[12] and the median for 51,620 native-born whites is 58.9.[13] Doctor Doll examined with Alpha 514 children from the eight grades of the Franklin School in Trenton, N. J., and found as follows:[14]

Median Age	9.5	10.5	11.5	12.5	13.5	14.5	15.5
Median Score	26	37	61	81	85	80	88
Number Cases	29	38	63	96	139	75	22

In this school the eleven-year-olds equal the army score, and again it seems that even with many allowances the army would not average higher than 13. It must be remembered in connection with the above figures that the medians for the upper ages are undoubtedly low, because a number of the more intelligent twelve- and thirteen-year-olds have gone on into the high school, while most of the fifteen-year-olds and many of the fourteen-year-olds left in the grades are mentally inferior.

A group of rural New York children selected from the fourth grade through high school and ranging in age from 8 to 21 gave the following scores:[15]

Age	8	9	10	11	12	13	14	15	16
Alpha Score	35	35	41	46	56	81	86	95	99

This gives us a more accurate idea of the scores for the upper ages, and again indicates that the average soldier was certainly not much above 13 years old.

If we turn to high-school results, we find the following, reported by Sylvester and Dobbs.[16]

Class	Madison		Rockford		Sioux City	
	Median Score	Average Age	Median Score	Average Age	Median Score	Average Age
Senior	121.2	17.3	123.8	17.5	137.6	....
Junior	122.4	16.6	119.7	16.4	130.2	16.6
Sophomore	109.7	15.8	107.0	15.3	122.1	16.0
Freshman	96.0	14.8	94.0	14.5	107.9	14.8

[12]Army report, Table 183, p. 666. This table contains a larger number of men than any other I can find.

[13]*Ibid.*, Table 309, p. 767.

[14]Doll, E. A., "The Average Mental Age of Adults," *Journal of Applied Psychology*, 3:317-28, December, 1919.

[15]Lufkin, H. M., "Report of the use of the Army Alpha test in rural schools," *School and Society*, 13: 27-30. January 1, 1921.

[16]Madsen, I. N., and Sylvester, R. H., "High school students' intelligence ratings according to the Army Alpha test," *School and Society*, 10: 407-10, October 4, 1919.

We must not forget in connection with these figures that the high-school pupils are a select group, but it does not seem that selection alone can account for the fact that the scores of the fourteen-year-olds are 50 percent and those of the sixteen-year-olds more than 100 percent above the average score of the army.

A valuable contribution in this field has been made by Kohs and Proctor.[17] They report a set of Alpha norms based upon the examination of several thousand California school children. Of this group they say, "The children tested were drawn from all types of communities, and represented every variety of social status." That part of the table which presents scores for twelve- and thirteen-year-olds is here reproduced:

Alpha Score	Mental Age Norms	
	Kohs and Proctor	Army
45	12–0	13–0
50	12–4	13–3
55	12–8	13–6
60	13–0	14–0
65	13–4	14–3
70	13–8	14–6

50.1 Median of 653 group
58.9 " 51,620 native whites
60.5 " 67,254 literate whites

These results indicate that the army mental age norms were somewhat high, and that the average man may have been slightly *under* thirteen mentally.

So the evidence piles up. The writer, however, does not wish to convey, through this article, his conviction that we should, without further investigation and analysis, adopt 13 or 14 instead of 16 as the average mental age for adults. His purpose is to set forth the fallacies in the arguments now generally advanced to invalidate the army results, in order that the question may be kept open and studied scientifically. If we are to use the results already in hand, the crux of the matter lies in the answer to the question: To what extent was the army typical of the

[17] Proctor, W. M., "*Psychological tests and guidance of high-school pupils,*" Journal of Educational Research Monograph No. 1, June 1921.

population at large? An analysis of some of the evidence has been presented in this article. Much further light can be obtained by the study of the psychological records now on file at the National Research Council. The study might well begin with the examination of the records of the men who were only twenty-one or twenty-two years old at the time of their entrance to the army. In this group the exemption for marriage and occupation would probably be low, and we would be rid of one troublesome factor. The examining of school children should continue, with care that the examinations are given far enough down in the grades, and far enough up into the high school to include all the children of ages 12, 13, 14, and 15, thus avoiding any factors of selection.

ATTEMPTS AT TEST VALIDATION

RAYMOND FRANZEN
Director of Research, Des Moines Public Schools

Measurement in education has passed its infancy. We no longer need to plead for recognition of the cardinal importance of finding exact indices of the quality which we are describing. Our attention has been diverted from the question of whether it shall be done, to the question of how it shall best be done. The increasing popularity of tests promises an early verdict from the school men of America regarding the practical consequences and immediate advantages of such measurements. This verdict will be transmitted to us in terms of the tests that are used, not in terms of the tests that could have been used. It is, therefore, of utmost importance that the tests be more carefully constructed and fairly evaluated.

Administration is not scientific in its choice of tests. Geographical preferences, advertisement, and the authority and prestige of the author are prominent elements in evaluation. Administrative exigencies, such as the price of a test, the time it takes to give it, and the convenience of its scoring facilities, will provide a sale for tests whether they have other virtues or not. Such considerations should, of course, be entirely secondary, since price, time, and convenience can only be of value if the test renders the service that is needed; and exact statistical virtue must form the basis of sound judgment. Cost should be considered only *in relation* to value.

Chief among the fundamental virtues which determine the value of a test are, first, that the test shall measure what it purports to measure, and second, that it shall be reliable. It is very important that a test have objectivity, that it be scaled, that it have usable standards, and that it exhibit certain other secondary characteristics; but if it is to be a test at all, it is essential that it tell the story which it pretends to tell, and that it tell this same story upon similar, successive occasions.

An attempt is here made to determine in respect to entering high-school pupils, the degree to which our so-called intelligence tests measure what they purport to measure. In this connection I may say that I believe the effect on measurement would be salutary if in naming our tests we used less readily such a general term as "intelligence." We might well substitute for it some expression descriptive of a feature of the criterion against which, in order to prove its usefulness, the test is checked. Thus a kindergarten test which attempts to predict future success in first grade, and whose value is judged upon a basis of the correlation of the test given in the kindergarten with future success in first grade, should be called "A Kindergarten Test of First Grade Readiness," rather than "An Intelligence Test."

It is very difficult to decide what elements of school life to use in validating our "intelligence tests." It is quite within the range of possibilities that the various authors of tests differ radically in their conception of what minutiae of behavior combine to form the power of adaptation which by their tests they are attempting to measure. It is, therefore, very difficult to choose criteria which will tell us how nearly the tests do the job. A more intricate and refined evaluation of any one test is possible in various ways, but a wholesale determination of relative value means the choice of some easily obtainable criteria. An experimental criterion including a careful observation and accurate clinical description of children is impracticable. The use of an individual Binet test is not essential for this purpose since the average of a dozen group tests would in itself probably be as good a measure of the intelligence of children as the results of a Binet test. The unreliability of marks given by teachers, and of the judgments of the intelligence of children by teachers, makes these unsatisfactory media for comparison. Age, used negatively, though it is an indication of how long it has taken the children to get as far as they are, is insufficient, because children are promoted for reasons other than readiness to go on. The use of the totals of all the tests is to some extent reasoning in a circle, though nearly everyone would grant that

the average of twelve or fifteen tests of power of "adaptation to new material" is a fairly accurate indication of what is usually meant by intelligence. We should be willing at least to admit that a test which correlates positively and high with the total of a number of tests is a better instrument than one which correlates either postively and low or negatively—though this might mean that the test in question measured something different. However, if we do not know what this "something different" is, this is an almost useless reservation. The intercorrelations and the partial correlations of a test with the total, irrespective of the others, is another factor which contributes to evaluation, though the use of partials would logically follow the determination of a very acceptable criterion so that we might then select the best team of tests to do the work which we have defined in this acceptable criterion.

The use of these elements as criteria will establish in a rough manner the efficacy of the tests. In this investigation we have examined fourteen intelligence tests with respect to such criteria. Table I indicates certain correlation data. The spread of the data is not great; largely because the material concerns but one high-school class. The average age of the fifty-seven pupils who took all these tests was 168.82 months, and the standard deviation of their ages was 10.73 months. This point —namely, the variation in age—is important. We do not so much wish to find how well a test distinguishes a normal ten-year-old from a normal thirteen-year-old as to find out how well it will distinguish a stupid thirteen-year-old from a bright thirteen-year-old. We need an instrument which will distinguish neural quality when neural change due to age is constant.

Some of the criteria above mentioned have been used here in spite of their insufficiency in order to make crude distinctions between our better and our worse media of measurement of intelligence, leaving for a more detailed and intimate investigation the discrimination between the better ones, and the determination of what it is that the others test. These data are at best a mere indication of even this crude distinction. As I

TABLE I.[c] EACH TEST CORRELATED WITH THE TOTAL AND WITH EACH OF THE OTHER TESTS

	Totals of all tests except the Wylie	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. National-A	.93														
2. Terman-A	.92	.85													
3. Otis	.92	.93	.86												
4. Haggerty	.91	.83	.85	.78											
5. Illinois General	.90	.85	.83	.90	.81										
6. Pressey-Survey	.87	.81	.82	.37	.76	.78									
7. Mentimeter	.81	.72	.89	.73	.82	.67	.69								
8. Thorndike Reading[a]	.81	.77	.72	.83	.77	.84	.63	.44							
9. National-B	.80	.76	.76	.73	.73	.80	.61	.60	.51						
10. Dearborn-1	.74	.64	.68	.60	.69	.52	.65	.64	.37	.48					
11. Dearborn-2	.72	.67	.70	.55	.66	.55	.64	.63	.39	.59	.60				
12. Pressey Cross-outs[a]	.70	.69	.58	.74	.66	.66	.50	.58	.48	.63	.56	.43			
13. Wylie[b]	.64	.55	.67	.65	.64	.60	.59	.54	.63	.48	.57	.32	.37		
14. Myers	.63	.53	.57	.56	.49	.49	.54	.57	.25	.39	.54	.36	.35	.34	
Standard Deviation	15.76	16.65	28.66	24.21	18.43	20.25	15.46	11.53	.45	14.25	12.28	12.47	11.02	18.35	15.17
Mean	143.60	143.70	113.53	136.04	122.61	106.95	139.91	103.28	7.24	146.09	64.46	64.84	73.67	85.00	57.30
Coefficient of Variability	0.11	0.11	0.25	0.18	0.15	0.19	0.11	0.11	0.06	0.10	0.19	0.19	0.15	0.22	0.26

[a]Thorndike Reading and Pressey Cross-outs are 20-minute tests.
[b]Wylie A, B and C is a 15-minute test.
[c]There were 200 cases to begin with, but N of these correlations is 57—being the number that took all the tests.

have already stated, there are only fifty-seven cases involved, but the number of correlations used, the interpretations of which supplement each other, makes this a less serious objection. We started with a whole freshman class in one high school, about two hundred, but before the administration of all fourteen tests was completed the number taking all examinations was reduced to fifty-seven.

Table I gives the correlation of each test[1] with the total (divided by ten for ease of statistical computation), and the intercorrelations of all the tests. The total was obtained by adding the scores of all of the tests except the Wylie. Scores on this test had not yet been tabulated when the total was obtained. Each test was roughly equated with the others in variability before the addition. Table I includes for each test the standard deviation, mean, and coefficient of variability (standard deviation divided by mean).

Table II gives the correlations of each test with the total, and the intercorrelations of all tests, with reading ability, as measured by Thorndike Alpha-2, rendered constant.

Table III gives the average of the correlations of each test with the other thirteen. This table is a rough check on the first column of Table I.

Table V gives the intercorrelations of the elements used as marks received the first semester in high school. Column 4 of the same table gives the correlations with the average of teachers' judgments of the child's power to adapt himself to his environment. These judgments were rendered after at least half a semester's acquaintance with the children. Column 6 of Table IV gives the correlations with age at high school entrance, or the length of time it took the children to get to high school. Naturally these correlations are negative and the higher the negative "*r*" is, the better the test is in that respect.

Table V gives the intercorrelations of the elements used as criteria. As to the method of obtaining the teachers' judg-

[1]The Dearborn examination for higher grades was divided into two parts since it was twice as long as an ordinary test. The first half was called Dearborn-1 and the second half was called Dearborn-2.

TABLE II. INTERCORRELATIONS OF TESTS WITH THORNDIKE READING RENDERED CONSTANT ($N = 57$)

	Totals	1	2	3	4	5	6	7	8	9	10	11	12
1. Mentimeter	.87												
2. Terman	.85	.92											
3. National-A	.84	.66	.68										
4. Dearborn-1	.81	.57	.64	.60									
5. Pressey-Survey	.80	.59	.70	.65	.58								
6. Haggerty	.78	.83	.68	.59	.68	.55							
7. Otis	.78	.72	.67	.81	.56	— .35	.39						
8. National-B	.78	.49	.66	.67	.36	.44	.62	.65					
9. Myers	.77	.53	.58	.55	.50	.51	.48	.65	.31				
10. Dearborn-2	.75	.55	.67	.63	.53	.55	.61	.44	.49	.29			
11. Illinois	.71	.61	.62	.57	.42	.60	.46	.67	.80	.54	.44		
12. Cross-outs	.61	.47	.38	.57	.46	.29	.52	.69	.51	.27	.30	.54	
13. Wylie	.24	.38	.42	.12	.47	.32	.31	.30	.24	.24	.10	.17	.10

ments regarding the ability of children to adapt themselves to the environment the following instructions[2] to the teachers were used:

"I desire a special report upon each of the pupils whose names are upon the list read in teachers' meeting. This report is to be made up by each teacher who has had the pupil in class for more than one-half semester. Use the slips of paper furnished you, making out a separate slip for each pupil. Write his name upon the slip and under it your score according to the following directions:

"Using a scale from zero to ten, indicate your estimate of this pupil's native ability, that is, his power to adapt himself to his environment. This may be interpreted to mean his natural ability to solve a new problem without assistance. Try to eliminate from your estimate all acquired characteristics, in other words, pay no attention to such questions as to whether he is lazy, is poorly prepared for his work, is timid, has bad habits of study, etc. A score of zero means that the pupil has no power of adaptation to the circumstances which surround him in your class.

A score of ten (10) would mean that this power was as nearly perfect as you would consider possible to a child of his age, or at least just as good as that of the brightest child you have ever known."

Some of the tests show a very good rank with respect to each of these various ways of viewing their value. The Terman for instance, is rather high in each of the lists. The Illinois is high in each except one, Table II, which means that it is a very good medium of prediction, but that it measures reading as well as intelligence (as far as the Thorndike test measures reading). The Thorndike Reading, though it is only a twenty-minute test, stands high in each of the tables, except Table II, in which it is of course, not included. The Otis, the Haggerty, and the National tests give a fairly good account of themselves. The Mentimeter has the highest correlation with the totals of all other tests, when reading is rendered constant. This would justify us in saying that if we wish to use a non-verbal test, we should use the Mentimeter. The Wylie stands very high in the tables, in consideration of the fact that it represents only fifteen minutes of testing. Its high negative correlation with age indicates that it measures position within an age group better

[2]These instructions were prepared by Mr. A. W. Merrill, the principal.

than growth itself. The presence of this desirable feature in the Wylie is verified by other data showing a wide spread within age groups. The standing in the tables of the Myers and the two Dearborn tests indicates the comparative inferiority of these tests in terms of criteria as here defined. The two Pressey tests vary too much in the different tables for us to risk a judgment.

Table II can be usefully employed to give indications of the degree to which the prediction of a test is valuable, apart from the prediction which is rendered by reading. It shows association of the remnants when agreements in terms of reading are "ironed out." We are most interested in that kind of prediction, since school marks and teachers' judgments of intelligence are based largely on ability to read, and we want to know those who are capable but who cannot read. The Thorndike Reading test and the Illinois General Intelligence test correlate highest with teachers' judgments because they both involve reading to such a great extent. The Illinois General correlates so high with the reading test that we may say that it is a measure of reading.

The intercorrelations as they stand in Table I and as they stand in Table II, with the reading partialled out, are valuable to determine which tests to use together. Unless it is merely a more reliable result which we are seeking, we should not use tests together which are highly correlated. If the correlations presented in Table I are valid, Otis and National-A should not be used together. Terman, Thorndike Reading, and Cross-outs or Wylie would be a good team, since each is of some value according to these criteria and their intercorrelations are low.

It is not out of place here, to add that the size of the correlations with these criterion elements is no reflection on the tests, but rather on the elements themselves. Even if we were to admit that the school marks are a desired goal and even if these marks were reliable enough to yield a high correlation, we should still be rendering a better prediction of this kind of success by our best intelligence tests than expert corn judges render of the yield of corn when they judge it at our Iowa fairs.

TABLE III.—AVERAGE OF CORRELATION WITH THE OTHER 13 TESTS

Terman	.75
National-A	.74
Haggerty	.73
Illinois	.72
Otis	.71
Mentimeter	.66
Pressey-Survey	.65
National-B	.62
Thorndike Reading	.59
Dearborn-1	.58
Cross-outs	.56
Dearborn-2	.55
Wylie	.53
Myers	.46

TABLE IV.—CORRELATION OF EACH TEST WITH CERTAIN CRITERIA

Test	With average of H. S. marks the first semester		With average of teachers' judgments of "power to adapt"		With age	
	r	Rank	*r*	Rank	*r*	Rank
1	*2*	*3*	*4*	*5*	*6*	*7*
Thorndike Reading	.51	1	.67	1	−.42	4
Illinois	.42	2	.62	2	−.47	2
National-B	.41	3	.55	3	−.33	7
Otis	.31	4	.52	4	−.45	3
Cross-outs	.30	6	.41	13	−.29	10
Terman	.30	6	.51	5.5	−.39	5
Wylie	.30	6	.48	8.5	−.49	1
Haggerty	.29	9	.50	7	−.32	8
Dearborn-1	.29	9	.48	8.5	−.24	12
National-A	.29	9	.47	10	−.30	9
Mentimeter	.26	11	.45	11	−.27	11
Dearborn-2	.24	12	.42	12	−.12	14
Pressey-Survey	.22	13	.51	5.5	−.38	6
Myers	.12	14	.28	14	−.14	13

TABLE V.—INTERCORRELATIONS OF THE ELEMENTS USED AS CRITERIA. AGE IS USED NEGATIVELY

	Marks	Judgments	Totals
Judgments	.82		
Totals	.34	.58	
Age	−.37	−.48	−.40

The correlations of four judges' scores of ears of corn with the yield of those ears were all about 0.20.[3]

The intercorrelations of the totals with the other elements of the criteria given in Table V indicate the limitations of these other elements. For example, since the average of teachers' judgments correlates only 0.58 with the sum of 13 tests, there is undoubtedly something wrong with the judgments. Part of their insufficiency as validation criteria is indicated by their high correlation with the pupils' marks. The marks are of June, 1921. The judgments are of December, 1921. They correlate 0.82. And still the judgments correlate only 0.58 with the totals of 13 tests. This means that the teachers' estimates of intelligence as defined in the instructions were very much more like their estimates of the product attained in their courses than they were like actual manipulation of new material by those children. Besides this, as the teachers' judgments are one element in these correlations and their reliability is very low, we cannot expect very high r's. A correlation coefficient is limited in size by the reliabilities of its elements. The correlation of any test with others cannot go higher than the square root of its correlation with itself.

These tables attempt to answer three questions in a limited and crude manner.

(1) Which tests are most likely to be good for all-around use?
The answer obtained from these data is Terman, Alpha-2 Reading, Illinois, the Nationals, Otis, Haggerty, and if a cheap and short test is wanted, the Wylie Opposites.

(2) Which tests are best to use if intelligence, irrespective of reading, is the objective?
The answer of these data is Mentimeter, Terman and National-A.

(3) Which tests should be used together and which test the same qualities?
The answer is to be found in Tables I and II.

[3]These data provided by H. A. Wallace of Des Moines.

More extensive and more intensive investigations of this nature will yield us fairly accurate determinations of the relative value of intelligence tests. In addition we must carefully demonstrate the quality of the quantitative messages that our product tests transmit. Here our methods are more clearly outlined. It is even possible in some cases to decide by analysis that a test measures some definite quality though the limits of such a method of validation are very restricted. Arithmetic tests do test arithmetic on the face of it, though it is instructive to compare Woody, Woody-McCall, and Courtis; and then, bearing in mind their differences, to hear teachers and principals talk of arithmetic scores, not "Courtis" scores or "Woody" scores.

Three of our well-known reading tests correlated as follows in one class of 40 children:[4]

	Thorndike-McCall	Monroe
Monroe	.26	....
Burgess	.19	.48

Although much more work with these tests is necessary to determine their true correlation, it is obvious from the data at hand that we need to investigate our reading tests. In a recent number of the Journal of Educational Research Courtis says:

> "The child who makes a high score in my test and a low score in McCall's may need to be given a series of life experiences rather than a series of reading experiences. To call McCall's a reading test is to imply that the child's deficiency can be made up by practice in reading. This is not true."[5]

I do not know whether this criticism is meant to extend to the Thorndike Alpha-2 or not. The type of the test is exactly the same. But I do know that a child's deficiency in the Alpha-2 can be made up by practice in reading. In an investigation at Garden City, New York, the correlations and differences between Intelligence Quotients and Reading Quotients

[4] These are correlations obtained by Miss Minnie Thurber of the Public Schools of Los Angeles, California.

[5] Courtis, S. A. "The analysis of reading ability." *Journal of Educational Research*, 4: 292, November, 1921.

(Thorndike Reading Age, divided by Chronological Age) all showed increasing correspondence of Binet scores and Alpha-2 scores after reclassification and training in *reading*, not life experiences.

This is one important way of validating a test. Does the correlation of the test with mental ability approach unity when training in the subject is given? If so, the product test and the mental test used are adequate. Woody-McCall Arithmetic, Thorndike Reading, and Stanford Revision Binet do test what they purport to test by this standard.

May Ayres Burgess in the 20th Yearbook of the National Society for the Study of Education says:

> "Since TB is the usual abbreviation for tuberculosis, print the two initials closely contiguous to this celebrated emblem of the American Tuberculosis Association. In doing so, be careful to select such locations as will result in the separation of the initials by the intervention of the emblem between them."
>
> Third-grade children, Italians, on the lower east-side of New York City, who were counted by their teacher as exceptionally poor in reading, and were not themselves able to use English fluently, read that paragraph, re-read it, and finally with shining eyes grasped their pencils, and printed T on one side and B on the other side, closely crowded against the central cross. They could not have defined the words, but they could discover what the paragraph meant if given long enough time to puzzle it out. The same result occurred over and over, with other forms of test material, even such as, under the picture of the veiled lady, requested the children to "Perpetrate an incision in the facial vesture immediately anterior to the buccal orifice through which the fastidious female may procure aliment."[6]

But when such a picture as this is put before third-grade children without any reading matter and the examiner merely says, "Do something to the picture," many of the children do just this.

If Picture Supplement-1 is given orally (that is, if only the pictures are put before the children and the instructions are given orally), and if Picture Supplement-2 is afterwards given according to instructions, except that plenty of time is allowed,

---

[6]Burgess, May Ayres. "Controlling factors in the measurement of silent reading." *Twentieth Yearbook of National Society for Study of Education*, Part II: pp. 29-30.

the correspondence of the results is marked. Third-grade children fail both ways; sixth-grade children get very nearly perfect scores with little spread in both ways of giving. Then reading has not much to do with it. In the lower grades, it is a test for following directions, and in the upper grades it is a test of speed of doing what one is asked to do, orally or in print.

Miss Kate Kelley, the primary supervisor of Des Moines, and I are constructing a reading test for I-B children, in which the first step in the process of test building is the administration of each proposed element orally, and the inclusion of only those that *practically all children get correct when given orally.* Then, all the difficulty is reading difficulty. It is impossible to guess which elements will fall low when given orally. Who could predict that when orally presented about 90 percent of first-grade children would say that "The sun sets in the morning" is true, or that "Boys and girls never sing" is true? Still this is a reliable determination, whereas seemingly harder sentences are correctly marked by about 90 percent of first-grade children.

When all sentences included are correctly marked by at least 90 percent of the children when given orally and from 50 to 100 percent when given in print, then it is fair to attribute the added difficulty to reading. This is a second possible method of validation.

Pressey[7] has suggested that we use the lack of overlapping of two distinct groups as a measure of validity. For example, when National-A is given to all children in high school, who received two or more of the lowest marks given, and to those who received two or more of the highest marks given, the overlapping is small. Only one or two cases out of 100 of the group marked high, fall below the average score of the group marked low, and only one or two out of 100 of the group marked low, fall above the average score of the group marked

[7]Pressey, S. L. "Suggestions looking toward a fundamental revision of current statistical procedure, as applied to tests." *Psychological Review*, 27: 466-472. November, 1920.

high. A study of these children demonstrates why. This is a third possible process of evaluation.

Teachers' judgment may be used to better advantage than merely as a measure against which to correlate test marks. For example, as a partial validation of the Park-Franzen test for readiness to do first-grade work, we had each of the fifty kindergarten teachers at Des Moines judge this readiness of her children; and we then correlated the test results for each class against the judgments of the teacher of the class. Thus we obtained fifty correlation coefficients. Each of them, however, concerned a series of judgments by a single teacher. Then arose the question of the ability of these teachers to make correct judgments; and, as the best answer to the question, Miss Park ranked them according to her estimate of their ability in this respect. If the correlation coefficients between test results and teachers' judgments were highest for the classes having teachers of best judgment and lowest for the classes having teachers of poorest judgment—in other words if the coefficients showed a high correlation with the teachers' ability to judge—this fact would be a favorable indication for the test. In other words, if there is a direct relation between a teacher's ability to judge, according to her supervisor, and the size of the correlation of her judgments with the test, it is a favorable comment on the test.

The intercorrelations with other tests is always a valuable determination; and whenever a criterion built up by careful clinical observation and repeated individual examination is possible, this is the preferable technique.

We are facing an era of selection of tests. We want fewer tests for measuring achievement in a given trait and better construction of such tests.

Editorials

REACTION

In the world of force and energy, there is a law that to every action there is an equal and opposite reaction. Students of history have pointed out that a similar law seems to operate in the growth of men and institutions.

In both plants and animals and in man himself, growth is neither uniform nor continuous. Periods of rapid change are followed by relatively longer periods in which very little growth is noticeable. So it is with national life and with social readjustments. In the few short years of the world war more new ideas were spread broadcast among all peoples than in any previous period in recorded history. It is inevitable that the reaction of the world to these ideas should be correspondingly profound; and it will be quite normal if, when the social order shall have stabilized on the new level, there should follow a long period of peace and prosperity.

Reaction, however, has sometimes a more sinister meaning. It has happened in human history that changes which have cost much in effort and blood have been completely wiped out by the reaction which followed so that the old order was not only reinstated but reinforced, and the coming of progress correspondingly delayed. Not every revolution against a tyrannical government has resulted in greater freedom for the revolutionists. More than one king has been able to stage a "come back." So with many minor details of progress. More than once the first experimental trial of a new procedure has lead to a reaction which wiped out every vestige of the new idea for at least a generation.

It is one of the outstanding characteristics of the times in which we live that the rate of progress has been enormously

increased, and at the same time the periods of reaction have been correspondingly reduced. Today change, progress is almost continuous. The most radical inventions, the most far-reaching changes hold the public mind only for a few short days. Their good points are quickly realized and they themselves are at once incorporated into our daily living. Such reaction as there is, is largely regret by the older generation for the passing of some phase of social custom made superfluous by the new, or objection to some new custom in which they do not share. Is a reaction against the telephone thinkable? Is it probable that the use of automobiles will some day be forbidden and the day of the horse-drawn vehicle restored? How long will the aeroplane and wireless remain mere marvels? Even now they are being put to commercial use, although such use involves a change of social relations and will result in social readjustments so profound that a few generations ago they could not have been brought about except by revolution and bloodshed.

The secret of the assimilative powers of the present generation is education. The invention of printing made universal education possible. Education in turn made possible direct appeal to individual thought and reason. Today publicity for each new invention, for each social change accomplishes in a few short days what formerly it required several generations to bring about. Where there is an efficient "selling campaign" which reaches all the individuals involved, there is no possibility of reaction.

For education itself this means that progress is dependent upon the organization and maintenance of effective forms of training for teachers in service and equally effective forms of letting parents know the reasons for the changes made. Changes in social life compel changes in subject-matter and in methods of teaching; and the education which the school provides is not permitted to lag far behind the education needed for the social life of the day. Society cannot wait for the slow processes of natural evolution to replace old teachers with the new. On the other hand, to make changes which are not understood by either the teaching force or the public is to invite reaction.

A single illustration will prove the point. Twelve years ago the first reliable instrument for the scientific measurement of educational products was invented. Ten years ago standard tests were used for the first time in a great city survey. Now tests and scales are to education what a ruler is to a carpenter, what a thermometer is to a physician. Their value has been established beyond all question. The contributions to education made by measurement have been many and far-reaching. The teacher who does not measure, who does not base her thinking upon the evidence obtained from scientific experimentation bears about the same relation to the one who does that the old country practitioner bears to the modern diagnostician. Today the teachers who are being graduated from our teacher-training institutions are not even conscious that they are measuring. Measurement has become an integral part of teaching and is accepted by the rising generation precisely as they accept the telephone and the automobile.

But what about the teachers in service and the general public? Rumblings of reaction are heard from time to time and progress is delayed. Fortunately, very fortunately, the teachers themselves have realized the situation and have crowded into summer schools and regular sessions of our universities and colleges to get for themselves the newer ideas and methods. Fortunately, the children carry word from school to home and serve to educate the parents. Fortunately, also, the children have accepted tests and realized their value. Thus it has come about that a great change has been successfully effected in education, and the danger of disastrous reaction avoided.

That school authorities can always hope to be equally fortunate by chance is improbable. Both the public and the profession must come to see that there can be no stable progress, no freedom from blighting reaction without deliberate, well-planned publicity, mutual understanding between the home and the school, and adequate provision in time and opportunity for the continued training of the teachers in service.

S. A. C.

Reviews and Abstracts

STOUT, JOHN ELBERT. *The development of high-school curricula in the North Central States from 1860 to 1918.* (Supplementary Educational Monographs, Vol. III, No. 3.) Chicago: University of Chicago Press, 1921. 322 pp.

The author first reviews the earlier high-school movement of the period previous to 1860. He shows that the curricula of the high schools changed slowly during that period. The aim of the Boston High School was for general education rather than preparation for college, while that of other schools was twofold, thus determining the type of school that has now become prevalent in the United States. The curricula of this earlier period covered a wide range of subjects and were not uniform although the differences were not striking.

For convenience the following period of 40 years is divided into eight five-year periods. A series of tables is presented made up of a study of the curricula of an increasing number of schools for each period. Little variation is found in the content of these curricula. Toward the close of the period some courses in commercial subjects appeared, but not to a sufficient degree really to characterize a curriculum. The same was true to a still less extent of normal courses for teachers.

The dominant course took the form of preparation for college with certain parallel elective courses in which some one subject, as science, might be emphasized. But there was no uniformity practiced aside from college preparation. The arrangement of prescribed and elective courses made its appearance more recently. The non-preparatory course with specific aim came very slowly.

The order in which subjects appeared in courses was practically the same as that still observable with respect to the older, so called, academic courses. Up to the close of the nineteenth century the prevailing custom was to divide the year into three terms. The semester division is a development of the last twenty years.

Chapters VI, VII, and VIII are devoted respectively to a discussion of "Constants and Variables," "Time Devoted to Fields and Subjects," and "Influence of Size and Location of Schools."

Part II treats of conditions and changes in subject-matter. Especial attention is here given to English, Sciences, the Social Studies, and Foreign Language and Miscellaneous Subjects. To students of the present day these studies are very enlightening as indicating the courses of work and the trend of development in the fundamental fields of high-school instruction.

Part III deals with "More Recent Developments," in which the writer clearly analyzes the general character and significance of the relatively rapid change that has taken place during the last fifteen years.

On the whole this is a valuable contribution to historical material in secondary education and the work shows a thoroughness and insight that is perhaps too exceptional in the subject-matter of many educational publications of the day.

H. A. HOLLISTER

University of Illinois

THORNDIKE, EDWARD L. *The psychology of arithmetic.* New York: The Macmillan Company, 1922. xvi and 314 pp.

"Satisfyingness in results" is a chief factor in fixing bonds. What attachments one forms for his old friend "Psychology," and what pride in the present dynamic subject, when he reads Thorndike's "Psychology of Arithmetic!" Professor Quack and Mr. Charlatan may tell us (free?) how "psychology" will keep us young and beautiful, and how to hold our husband's love; but Thorndike for little or nothing lays bare the inner workings of the teaching and learning process that thousands of teachers may become more skilled and millions of children more efficient. The "medicine man" in psychology may obtain his knowledge and inspiration in a week's correspondence course, but twelve years ago the writer sat in Professor Thorndike's classes in the "Psychology of Elementary School Subjects" and heard this treatise on arithmetic "in the making." It has been "in the making" for many years since and the modesty of the author's claim is in significant contrast to the psychological bootleggers above, when in the preface it is stated: "So they (the lectures) are now published in spite of the author's desire to clarify and reinforce certain matters by further researches!"

Although not at present able to define in detail the formation of such bonds as are more generally stated as ability to read, spell, add, subtract, etc., "the psychology of school subjects begins where our common sense knowledge of these functions leaves off." In terms of problems the task, however, may and should include a study of the following topics:

1. What is the function—just what in detail is an understanding of decimal notation, the moral after-effects of literature, etc.?
2. How are degrees of ability or improvement in these functions measured? For example, what degree of spelling word difficulty or knowledge of a fraction should we attempt in grade IV?
3. What can be done toward securing better control of the particular situation—response connections?
   For example, what are the similarities or dissimilarities between the bonds formed in learning 7 and 4 are 11, 7 and 5 are 12, and 27 and 4 are 31, 27 and 5 are 32, etc.?
4. What sets of bonds produce most advantageously the desired changes in knowledge, skill, or powers? Should the pupil be taught that "0 and 0 are 0," "0 and 1 are 1," etc., or taught the habit of entering 0 to a column of zeros and neglecting the 0 in general in addition?

5. What degree of strength is most desirable in bonds at given stages of learning? Should certain denominate number tables be learned at their first appearance to be remembered a month, a year, or for the work at hand?
6. What is the best order to follow in fixing sets of bonds when change is desired? In early writing habits, should details of stroke precede or should children "write" words and sentences? Metric system bonds are usually found late in the course. Should they be formed early enough to facilitate knowledge of decimals?
7. What are the original tendencies and pre-school acquisitions upon which the school may base or counteract its connection forming?
8. What armament of satisfiers and annoyers, of positive and negative interests and motives, stands ready for use in the formation of uninteresting connections between black marks and meanings, etc. "School practice has tried, more or less at random, incentives and deterrents from quasi-physical pain to the most sentimental fondling, from sheer cajolery to philosophical argument, from appeals to assumed savage and primitive traits to appeals to interest in automobiles, flying machines, and wireless telegraphy. Can not psychology give some rules for guidance, or at least limit experimentation to its more hopeful fields?"
9. How do general laws of efficient learning apply in the case of each elementary school task? Self-competitive arithmetic drills have proved effective in interest and results. How may other arithmetical functions be made to conform?
10. How far are special abilities or difficulties inherited?

Such are the problems. The book discusses these with a wealth of concrete illustration in the field of arithmetic. The author has literally laid bare its beginnings and processes. At the close of the chapter on "The Nature of Arithmetical Abilities" he makes a significant comment: "What may be called a 'sociology' of arithmetic is very much needed to investigate this matter. For rare or difficult demands the elementary school should not prepare; there are too many other desirable abilities that it should improve. . . . It seems probable that the school training in arithmetic of the past has not given enough attention to perfecting the more elementary abilities. And we shall later find further evidence of this. On the other hand, the fact that people in general do not at present use a process may not mean that they ought not to use it. . . . When we have an adequate sociology of arithmetic, stating accurately first who should use each arithmetical ability and how often, we shall be able to define the task of the elementary school in this respect. For the present, we may proceed by common sense, guided by two limiting rules. The first is—'It is no more desirable for the elementary school to teach all the facts of arithmetic than to teach all the words in the English language, or all the topography of the globe, or all the details of human physiology.' The second is—'It is not desirable to eliminate any element of arithmetical training until you have something better to put in its place'."

The volume is replete with stimulating suggestions. Professor Thorndike has added a very concrete contribution to the better teaching of arithmetic

as well as to teaching in general. It may well be read by the classroom teacher.

CYRUS D. MEAD
University of California

KUHLMANN, F. *Handbook of mental tests.* Baltimore: Warwick and York, 1922. 208 pp.

This handbook presents a new revision and extension of the Binet-Simon intelligence scale. The revised scale is given to the public after seven years of continuous work on the part of the writer, Director of the Minnesota Research Bureau in Mental Deficiency, and a number of assistants. The chief emphasis of the book is directed toward the presentation of the system of mental tests; but many important items of information in regard to the general principles of the year scale and the conduct of a mental examination are included, and a table of intelligence quotients is appended. Since, valuable as these latter features are, the claim of the book upon public attention will be based primarily upon the contribution which the new revision makes to the field of intelligence measurement, the review is concerned primarily with a comparison of the revision with the Binet-Simon scale and with other revisions.

This comparison touches upon the following points: (1) total number of tests included in the scale; (2) number which are reproductions or modifications of Binet tests or tests of other revisions; (3) number shifted from the age-groups assigned in other scales; (4) number of new tests; (5) range of the scale; (6) general procedure in the administration of the tests, nature of the instructions to the subject, and method of scoring; and (7) number of subjects used in the standardization. The scales compared are the Binet-Simon scale, 1911 edition, the Kuhlmann 1912 revision, and the Stanford revision.

The new Kuhlmann revision comprises 129 tests (including repetitions), as compared with the 54 of the Binet-Simon 1911 edition, the 75 of the Kuhlmann 1912 revision, and the 90 of the Stanford revision. Of these, 37 are found in similar or identical form in the Binet 1911 edition, 61 in the earlier Kuhlmann revision, and 45 in the Stanford revision. Thirty-five tests are common to all four scales. Only 5 of the 45 tests common to the Stanford revision are found in neither of the other two scales compared. Sixteen tests are shifted from the position which they occupied in the Binet revision, 9 from that occupied in the earlier Kuhlmann revision, and 14 from the position assigned by Terman. The number of new tests is 60 if the first appearance of a test as well as each repetition is counted; or 23 if repetitions are not counted—approximately the same as those added in the Kuhlmann and Terman revisions considered.

When the 23 new tests are analyzed, it is found that a strikingly large proportion involve some use of geometrical forms, considered in detail, they require making simple learned responses, recognizing objects, naming from memory, tracing, dealing with geometrical forms, counting, following directions, imitating folding, giving word opposites, spelling words backwards, crossing out letters in a pied text, working arithmetical problems, and making

logical inferences. The range is from three months to fifteen years, as compared with a range from three years to adult in the Binet series, from three months to fifteen years in the earlier Kuhlmann revision, and from three years to superior adult in the Stanford.

In the general procedure in the administration of the tests, the instructions to the subject, and the scoring of the responses, the old tests appear to be patterned, almost without exception, directly after the earlier Kuhlmann revision, the text in instances being a verbatim reproduction. The resemblance to the text of the Stanford revision is also very striking, due to the common origin of the tests, although, as a rule, Terman's presentation is considerably more detailed. The new tests are very much like the older tests so far as their administration and the nature of the instructions are concerned; they differ markedly in the scoring, however, practically all those added for the years above two being scored in terms of the time required to perform the test and the number of errors made. The standardization of the scale is made as a result of tests with over 2000 normal children, in comparison with the 200 normal children used in the arrangement of tests for the Binet scale, the indefinite number employed in the earlier Kuhlmann revision (available data on the subject from other investigators having been included as well as that secured by Kuhlmann himself), and the approximately 1000 unselected children and 380 adults used by Terman.

In view of these comparisons, it is evident that the new revision is built primarily upon the Binet tests and the author's earlier revision, and in but small degree upon the Stanford. As compared with the first two scales, it is doubtless highly superior; as compared with the Stanford, it may or may not mark an advance.

Points which must arouse unfavorable criticism are the lack in objectivity of the infant tests and the inadequacy of their standardization, failure to include tests for the measurement of the intelligence of adults of superior mental capacity, the large quantity of test material required, and the elaborateness of the scoring formulae and the procedures for the computation of the mental age. Topics which will be subjects of controversy are the accuracy of the new positions assigned the tests shifted from the age-groups in which they were placed by other investigators, the validity of the new tests, the frequency with which the same test is used in different age-groups, the conception that the rate of mental growth decreases with increase in age, and the use of the results of examinations only of normal children in the establishment of norms. Admirable features of the scale are the simplification and improvement of the technique employed in the administration of some of the older tests, the elimination of some tests too largely dependent in their scoring upon the personal equation of the examiner, the introduction of others the scoring of which is almost wholly objective in character, the provision for the computation of a satisfactory mental age for those who pass more than two tests in the highest age-group, and the inclusion of a table of intelligence quotients.

Final evaluation will have to await wide use of the new scale by competent examiners and the determination of its validity in terms of its corre-

lations with other criteria of intelligence. It will also have to await the determination of the comparative validity of such remarkably simple and convenient scales as the Herring revision of the Binet-Simon tests, and the various group tests for the measurement of intelligence.

LAURA M. CHASSELL

Ohio State University

DEWEY, EVELYN. *The Dalton laboratory plan.* New York: E. P. Dutton and Company. 1922. 173 pp.

Miss Dewey who has for many months been in close touch with Miss Helen Parkhurst, the originator of the Dalton Laboratory Plan, describes this experiment in school organization.

The fullest establishment of Miss Parkhurst's ideas is in her own elementary school, the Children's University School, in New York City. There the children, beginning with the fourth grade, spend their mornings from 8:30 to 12 in "laboratory" work. The hour from twelve to one is occupied with class conferences and auditorium exercises. The afternoon is given over to work in special subjects. In the three and one-half hours of each morning a child is free to work on the subject or subjects which he wishes—and in whatever order he chooses for as long a time as he thinks best.

Two factors in the organization which make possible this freedom are the "subject laboratories" and the monthly "contracts" of work in all subjects. Each of the five academic subjects in the curriculum is to be found in its own laboratory—a room in which all the teaching and studying in that subject is done. When a child goes into a laboratory he finds there the subject teacher and numbers of other children all engaged in work on the same subject, though they may be doing any phase of it from fourth-grade to eighth-grade complexity. It is a rule of the laboratory that a child may cooperate with any others of his own grade who are in the laboratory, help younger children, or ask for assistance from older pupils. The social experience within the child's own group and with other grade-groups, Miss Parkhurst considers two of the important aspects of her "laboratory" plan.

At the beginning of each month assignments in the subject for each grade are posted in the laboratories. The child copies the assignments on his record card, and this constitutes his contract for the month. The assignments in the subjects are arranged in blocks which are equivalent to one or two day's work, so that the child may enter on his record card how much he has accomplished in his morning's work in a given laboratory.

Children develop various types of procedure. Some children prefer to finish all the assignments in one subject and then go on to another. Other children do a little on each subject each day. How they may apportion their time is a matter left entirely to the children's discretion. Practice in controlling the day's activities is the unique contribution which the laboratory plan makes to children's educational experience.

Miss Dewey's account of the plan includes the detailed arrangements for the teacher's supervision of the child, the keeping of records, the making of assignments, the nature of assignments, and the like.

To quote Miss Dewey in her summary of the book—"Complete experiences, individual work, and scope for interests are essential in both the Dalton plan and the project method; when either is explained in theoretical terms.... The project method has freed itself from one side of this conception, the school of classified and isolated facts. The Dalton plan frees itself from the other half, the school of piecemeal assignments, bells, and herd-learning and recitations. One contributes a new subject matter to meet the needs of modern life; the other a way to give children working conditions that accord with the discoveries of modern psychology."

HARRIET O'SHEA

Bryn Mawr College.

VOELKER, PAUL FREDERICK. *The function of ideals and attitudes in social education.* Columbia University, Contributions to Education No. 112. New York: Teachers College, Columbia University, 1921. 126 pp.

"You can not measure the human soul." A decade ago such was the slogan of the conservatives in their opposition to educational measurements. The slogan implied in particular that it is impossible to measure achievement or improvement in such school subjects as handwriting and arithmetic. But the events of the past ten years have shown that the conservatives were mistaken. Numerous tests are now available for use in practically all of the elementary-school subjects, and there are hundreds of practical school people who are ready to testify that the tests are worth while. Marked success has also been attained in the measurement of intelligence. Now Dr. Voelker comes with a book in which he undertakes to measure the affective side of mental life. The contribution is pioneer and even epoch-making in character.

The organization of the book is logical. The first chapter is entitled "The Perspective." The second chapter deals with the objectives of "Social Education." The third is devoted to the "Problem of Methods," the fourth to the "Problem of the Function of Ideals". The remainder of the treatise deals with the data obtained in an experimental manner.

In his second chapter the author defines his problem and marks out some of the chief objectives toward which social education should be directed. As such he specifically discusses the ideal of trust-worthiness, loyalty, social service, social sympathy, social conscience, social cooperation, social initiative, social justice, social control, tolerance, reverence, and faith. While some may prefer to state matters differently, all will agree that every effort to promote a clear conception of the objectives of social education is worth while. Definiteness of aim is nowhere more valuable than in the realm of morals.

Dr. Voelker's work would have been valuable even if he had stopped at this point. He goes further however. Having established his objectives he proceeds to a consideration of the problem of "Methods in Social Education." The essentials of such methods are stated in twelve assumptions which constitute the most complete summary of the sort which has yet appeared.

The author's most important contribution however, is in his last chapter in which he supports his theses by evidence, derived from actual experimentation. Having decided to make the ideal of trustworthiness the objective of

the training, two series of tests of trustworthiness were devised which are as nearly alike in character as possible but quite different in content. Series 1 was given to four groups of boys and the scores were obtained. Then two of the groups received training in trustworthiness by methods "such as are commonly used by the leaders of the Boy Scout Troops;" the other two constituting control groups. After seven weeks of training the second series of tests were given. The scores were tabulated and compared with those of the first test. The two experimental groups had improved 13.5 per cent and 9.3 percent respectively while both control groups had failed to improve.

The elements of the experiment which are both unique and significant are these. The author has shown quite clearly that it is possible to provide a type of training which will strengthen and enlarge the ideals of the pupil. He also shows that improvement obtained from training, which is general in character can spread into various specific cases in which a general ideal is supposed to function. Tests have been very ingeniously devised which measure such an abstract and intangible quality as trustworthiness. A parallel series was provided so that improvement in the quality could be measured. Thus the way has been broken for the measurement of the affective phases of mental life.

Future investigations will bring many refinements of technique. Better methods of scoring, more adequate criteria and greater reliability and validity of the measures will doubtless be secured but this study will remain conspicuous as one of the first successful efforts to measure abstract personal qualities of the affective type by objective standards.

W. J. Osburn

Department of Public Instruction, Madison, Wisconsin

Niemeyer, N. *Children and childhood.* New York: Oxford University Press, 1921. 206 pp.

The purpose of the author is to give a cross-section of the child at various stages from birth to sixteen years. The reader is not to consider any characteristic peculiar to one period, but at a particular age, some given characteristic may stand out more than at another. The life of child is divided into the following periods: 1 to 5, 5 to 7, 6½ to 8½, 9 to 11 or 12, 12 to 16. The first is characterized by the education of muscles and nerves. The predominant change is the establishment of nervous connections in the cerebral cortex. Lack of nervous connections accounts for slowness of reaction on the part of the very young. For developing these, the child needs freedom of movement, nourishment, large playthings, and small things to use. Proper brain connections should be established by the formation of habits. The diagrams used to illustrate this leave much to be desired in respect to accuracy.

When the child commences to classify objects, however crudely, he has passed the first stage of babyhood. Imagery and imagination now stand out prominently in his play. In these he differs from his younger or older brothers and sisters, for the younger ones enjoy play for sensory repetition, while the older ones desire more reality. Emotions of self and sex are also becoming evident. The writer agrees heartily with Miss Niemeyer in the fact

that children should be taught simply but accurately facts about sex. Evasions lead to false suppositions. This is also the period of instilling proper habits of discipline and thinking. Discipline should not be repressive.

The period of 6½ to 8½ years is one of transition. The child differentiates between the real and fantastic. Dullness and backwardness, if present, are more evident than previously. Between the age of 9 and 12, the child shows more stability, physically and mentally, than in babyhood. During this period especial attention should be given to the training of the memory. Miss Niemeyer also emphasizes the necessity of transfer of habits at this age by producing classroom situations more like those of the world, for only when the two situations are alike in one essential point is there transfer. She hastens to explain in a footnote that this really signifies no transfer. The child's thought and interest are self-centered. Attention to these, however, does not have to be created, but has to be trained.

The last period, or that of adolescence is marked by the arousal of sex- and self-emotions. Self-emotion leads to reflective thought and self-willed actions. During this age, the adolescent should be taught to decide his own problems of discipline. All of these factors properly guided should lead to the establishment of a strong character.

Miss Niemeyer's book is well worth the time of perusal of the mother interested in the psychology of her child or the student beginning the study of child life. It would serve as a good introduction to the possibilities of the subject. But for the more mature student, the book is scarcely thorough enough. One of the best features of the book is the original excerpts appended to illustrate various characteristics of the different ages of childhood. After reading this little volume one should understand children better in the light of their activities and conversation; and that, after all, was the main purpose of the author.

A. Sophie Rogers

Ohio State University

Bamberger, Florence Eilau. *The effect of the physical make-up of a book upon children's selection.* Baltimore: The John Hopkins Press, 1922. 161 pp.

The monograph presents the results of an experimental study of the factors which influence children's selection of books for reading. The desirability of such empirical approach is brought out for it develops that adults' judgments regarding the books which children will choose disagree markedly with the actual findings.

The material consisted of five familiar children's stories each in five editions. These are described in detail and apparently differed widely in physical properties—external and internal. Each story in its five editions constituted a separate experiment. The story was read by the experimenter up to a certain point of suspense and the children were asked if they would like to finish reading it themselves. The five editions containing that story were then placed upright on the chalk rail of the blackboard and a number written above each. The space error was obviated by altering the arrangement of the books

with the different groups of subjects. The children came forward one at a time, inspected the books without touching them and recorded on a slip of paper the number of the book in which they wished to finish the story. Each child was later given the five editions, was allowed to examine them freely and was then asked to indicate the one preferred. If this choice differed from his first ("inspection") choice he was asked the reason and his comments were noted. The experiments were performed with groups of about ten children at a time. The teachers acted as experimenters and recorded the original results. The subjects were pupils in the first three grades of five widely differing schools in New York and Maryland. Each series involved results from 76 to 260 subjects.

The results of the experiments are presented in great detail with tables and charts. The size of the book seems to be a significant factor in selection. This result is further corroborated by additional experiments in which all the books were covered with tan paper. This same experiment shows that the appearance of the cover also exerts an influence. Both the color of the cover and the nature of the title are involved. It is interesting to note, with the older children, the negative effect of the word "reader" in the title. The outstanding factor influencing choice, after thorough examination, is illustration. In the books that were chosen more than one-fifth of the time (chance expectation) the average amount of space devoted to illustrations is 45 percent while in the books that were chosen less than one-fifth of the time the average amount of space devoted to illustrations is 27 percent. There seems some evidence that illustrations portraying humor or action and having story telling qualities are preferred. Few lines to a page and wide margins seem desirable. The fact that in 45 percent of the cases the books were retained by the children after an opportunity to examine them, points to the benefits of open shelves in libraries. Differences in sex and school grade throughout the experiment are slight.

The monograph concludes with a score card for evaluating the various factors in the physical make-up of a book for primary children. Each of a large number of factors is weighted according to its actual frequency of appearance in some 2000 choices. Then fourteen of the books used are evaluated according to this scale. Their values range from 76 to 38 and correlate perfectly with their actual frequency of choice.

The method of the experiment seems sound and if strictly adhered to should yield results void of suggestion and space errors. The number of subjects seems sufficient to afford valid conclusions. The weakest part of the experiment in the reviewer's opinion is the material, involving as it does so many simultaneous variables.

The reviewer intends by no means to belittle the monograph. It is a distinct experimental contribution in a field that has not been hitherto approached in this manner. The method is novel, the subjects adequate and the presentation of results detailed and lucid. It points the way to further study of certain variables which did not come out clearly with the present technique but which are perhaps of considerable significance.

HAROLD E. BURTT

Ohio State University

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

Doctor John L. Stenquist, assistant director of the Bureau of Reference, Research, and Statistics at New York City, has been appointed Director of Educational Research for the public schools of Baltimore, Maryland.

Doctor Leo J. Brueckner, assistant dean of Detroit Teachers College has accepted an assistant professorship in the College of Education, University of Minnesota, beginning with the current semester.

Doctor Marion R. Trabue, head of the Bureau of Educational Service, Teachers College, Columbia University, has resigned that position to become Professor of School Administration in the University of North Carolina.

At Stanford University, Dr. Lewis M. Terman has been appointed head of the department of psychology to succeed Professor Frank Angell who retires at the end of the academic year 1921-1922. Other additions to the department include Dr. W. R. Miles, professor, and Dr. Calvin P. Stone, assistant professor.

Doctor William A. McCall, assistant professor of school administration, Teachers College, Columbia University, is spending the current year in China on invitation of the National Association for the Promotion of Education to assist in developing scientific education through the creating and adapting of tests, formulating experimental and statistical methodology, and advising about researches.

The Educational Research Department of the Ohio State Teachers' Association met at the Hotel Breakers, Cedar Point, Wednesday evening, June 28, 1922. Some months ago the members voted by mail to dispense with formal programs at the summer meetings but to hold a dinner, followed by an after-dinner address. The meeting was in charge of the vice-president, Doctor A. R. Mead of Ohio Wesleyan University. Doctor E. J. Ashbaugh of Ohio State University addressed the meeting upon the theme "Some Essentials in Supervision." The address was well presented and well received. The organization instructed the secretary, Superintendent F. J. Prout of Sandusky, to mimeo-

graph the address and mail copies to all public school superintendents in the state as well as to the members of the department. This was done within two days of the date of the meeting.

Our readers will be interested to know that the Charters Diagnostic Language and Grammar Tests will be made more useful this year than they have ever been before. We have all recognized the fundamental soundness of Doctor Charters' language and grammar studies—studies based on the errors which children make. Unfortunately the particular form of the tests has not hitherto exhibited the best technique, the result being that the tests have fallen short of the objectivity which modern standards require. The new forms of the tests will be ready for distribution in September. With the cooperation of a number of superintendents the new edition is being standardized so that tentative norms will also be available this fall. The tests will continue to be published by the Public School Publishing Company.

We recently received a brief letter which, with the omission of the name of the city, read as follows: "I am very happy to tell you that we have no supervising principal in ————. In my opinion that would be the greatest non-essential that could be added to the system." We naturally wonder concerning the value of this opinion. Has the writer scientifically evaluated the work of supervising principals? If so, and the results have shown them to be non-essential, then this is not an opinion but a statement of fact. On the other hand, if such an evaluation has not been made, from what source has the writer derived the right to express an opinion upon the subject?

The London Teacher for July 7, 1922, calls attention both by editorial comment and news item to an action taken by the education committee of the London County Council which will be of great interest to educators everywhere. The report contains a very significant statement, namely, "we have considered this question very carefully and are of the opinion that the children would not suffer if the classes mainly composed of children under five were taught by the same type of women as was employed for this purpose during the war." The type of teacher referred to was the emergency teacher who had received a three-months intensive course of preparation for teaching very young children. The argument for this action is an economic one, the difference in salary between a certified teacher and an uncertified one being about one hundred five pounds per year, or something less than five hundred dollars. An experimental arrangement with a group of one hundred young teachers is to be attempted. The question is raised editorially and we may well repeat it in this country as to whether it is economical for unprepared teachers to teach children of any age.

The following expression of attitude adopted by the New York Federation of Labor in its fifty-eighth annual convention this past June is worthy of the most careful consideration of school people everywhere.

"School statistics show that the great mass of children leave school between 14 and 16 years of age, and that the eighth grade or lower is the

scholarship attained by this group. Many parents and children regard graduation from the VIII-B grade as the end of education; and this impression is heightened by the elaborate public ceremonies attending graduation exercises when some overzealous trustee or other prominent citizen attempts to advise the graduates on the brilliant careers open to them. There should be no break in the continuity of advancement to the higher grades, and *graduation exercises should only be held on completion of the fourth year in high school.*"

The chairman of the committee on education, in forwarding this statement to the school people of the state, recommended the abolition of elaborate graduation exercises and the discouragement of the tendency toward elaborate apparel for graduates.

Rochester, New York, has been one of the leading cities of the United States in the establishment of junior high school. Four junior high schools are now in operation; and by 1925 it is expected that the entire city will be on a 6-3-3 basis. The development of the work has been gradual, and the Municipal University has devoted part of its energy to the preparation of teachers for this type of work. The salaries in junior high schools are the same as in senior high schools, and the teachers must have had two years of normal-school training, one year of college work, and three years of experience.

The first junior high school in the city was organized in 1915 in the Washington school district which has more than 90 percent foreign population. The average carry-over to the ninth year for five semesters previous to the opening of the junior high was 58.7 percent. The carry-over for the succeeding ten semesters averaged 85.8 percent. In comparison with this, the carry-over in the Madison school district, which is a representative American section, was 73.6 percent and 78.2 percent for the corresponding periods. Such information presents real evidence of the effect of the junior high.

Assistant Superintendent J. P. O'Hern, who has furnished us with these data, further states: "The junior high school is a great factor in Americanization work. In the Washington section the Jews would remain in No. 9 school, the Italians in Nos. 18 and 27, the Poles in 22, the Germans in 26, etc., until they graduated from the eighth grade. Now all the foreigners enter the junior high school before the law allows them to leave school. They touch elbows, become friends, and consequently get rid of their natural prejudices and hatreds. The junior high is democratic in offering facilities according to needs and abilities and in bringing pupils together at a time when friendships are formed and prejudices removed."

The following communication from Dr. Frank N. Freeman of the University of Chicago, concerning some of the work in the educational laboratory there, will be of interest to those who are interested in the development and use of refined technique in the evaluation of educational procedure.

"The investigations of the mental and physical development of children have been conducted for the most part by making measurements or tests of different children of different ages. The averages of the successive

ages have then been used to construct a curve of development which has been regarded as the norm. We have no assurance, however, that such a curve represents at all faithfully the course of development of individual children.

In fact the few studies in which repeated measurements have been made in succeeding years upon the same children have indicated that individuals deviate to an important degree from such averages. In the second place it has not been usual to make a variety of both physical and mental tests of the same children in order that the relationship between the different aspects of the child's development may be correlated. There have been, it is true, exceptions to both of these generalizations. It is desirable, however, that a much more comprehensive, systematic, and long-time study of the development of a group of children be made than has been made up to the present time. Such a study is being initiated through the cooperation of the educational laboratory and the laboratory schools in the University of Chicago. The plan contemplates keeping record of the physical measurements including X-ray photographs of the bones, making physical and psychophysical tests, and making a record of educational achievement and of changes in emotional attitudes on the same children for a period of ten years, or as long as the children remain in the schools. It will be possible to secure from time to time comparisons between the results of these various tests, but the fundamental object of the study will be to trace the various forms of development as they appear in the same individual children and to study their relationship to one another throughout the period of the investigation."

Inferences from Small Samples and Differences

In a recent article[1] in this Journal, Mr. J. W. Heckert has presented an excellent plan for determining the value of supervised study in English composition. The improvement of supervised over unsupervised groups is shown by the gains over preliminary test scores on the Thorndike-Hillegas Scale. Two other types of comparison are made by separating the groups into halves according to intelligence (National Intelligence Test), and according to initial test score. From all of these plans the writer finds differences in favor of the supervised group, and concludes that, "Supervised study in English composition is eminently worth while when teachers are able to direct the children's efforts intelligently."

While the plan of this study is admirable, the conclusions from the results obtained are open to question because of the small number of cases involved (8, 16, 17), and the slight differences in gain. According to Table I, the average gain for the supervised group is 2.6 percent. The question arises as to whether the difference of 4.4 percent is really significant or not. On this point the validity of the whole conclusion rests. The obvious answer is to determine the probable error of the difference by the formula

$$P.E._{a-b} = \sqrt{P.E._a^2 + P.E._b^2}$$

[1] Heckert, J. W., "The effects of supervised study in English composition," *Journal of Educational Research*, 5: 368-80, May, 1922.

The quantities $P.E._a$ and $P.E._b$ are obtained for the respective mean gains from the formula

$$P.E.\ \text{mean} = .67449\ \frac{S.D.\ \text{distribution}}{\sqrt{N}}$$

giving

$$7.0 \pm 1.5 \text{ and } 2.6 \pm 1.4$$

The probable error of the difference between the two means is then

$$P.E.\ \text{difference} = \sqrt{(1.5)^2 + (1.4)^2} = 2.1$$

Therefore difference $= 4.4 \pm 2.1$

By making use of Pearson's Tables of the Probability Integral we find that the odds are roughly 5 to 1 against these two means being identical, or merely random samples from the same population, or that the difference is really significant. As a matter of fact the probable errors for such small samples are even larger than those given by the above formulae, and the chances for significance of the differences even smaller. The usual rule is that the observed difference must be three or four times its probable error in order to be significant. This places the odds at roughly 100 to 1 against the difference being due to sampling.

Differences and their probable errors have been calculated for the remainder of the writer's results and are given in the table below. Inspection of this table shows that none of the differences are truly significant, the greatest odds in favor being roughly 12 to 1. These are hardly safe betting odds on which to base scientific conclusions. It is interesting to note, however, that the most reliable results revealed by the table are that the brighter pupils profit more by supervised than by unsupervised study, and that the initially poorer pupils on the test showed the greater improvement. This suggests that the bright lazy boy will profit under supervision, while the dull hard worker will not, a result often checked by experience.

MEAN GAINS AND DIFFERENCES FOR SUPERVISED AND UNSUPERVISED GROUPS

GROUP	SUPERVISED		UNSUPERVISED		DIFFERENCE	ODDS FOR SIGNIFICANCE
	Mean	S. D.	Mean	S. D.		
All	7.0±1.5	8.8	2.6±1.4	8.5	4.4±2.1	5:1
Brighter half	9.8±2.4	10.0	1.7±1.9	8.1	8.1±3.1	12:1
Duller half	4.3±1.5	6.4	2.7±2.2	9.3	1.6±2.7	.....
Better half	3.7±2.2	9.3	0.2±2.0	8.4	3.5±3.0	.....
Poorer half	10.4±1.6	6.8	4.1±1.8	8.1	5.8±2.4	9:1

It is to be hoped that the above experiment will be repeated with larger numbers, or adapted so as to reveal greater differences if they exist. If it seems inadvisable to use larger groups or different methods of supervision, a longer training period might yield significant results.

KARL J. HOLZINGER

University of Chicago

Intelligence Tests and College Success

During the first full week of the present academic year four different group-intelligence tests were administered to the one hundred new students entering Hood College for the first time. One test was given on each of four successive mornings. The four tests used were: Terman, Otis, Thurstone, and Rogers.

A study of the scores seems to justify two conclusions:

1. *That the four tests measure the same thing, whatever that may be.* In support of this first conclusion the following correlations are given (the method used is that of the percent of unlike-signed pairs):

	Coefficient of correlation	*P. E.*
Terman and Otis	.90	.01
Terman and Thurstone	.86	.02
Terman and Rogers	.61	.04
Otis and Thurstone	.90	.01
Otis and Rogers	.66	.04
Thurstone and Rogers	.73	.03

The following correlations were obtained between each test and the average of the four tests:

	Coefficient of correlation	*P. E.*
Terman	.93	.01
Otis	.97	.005
Thurstone	.96	.005
Rogers	.83	.02

2. *That the thing measured by these tests in this case was not the thing which determined success in college.* In other words, the scores derived from these tests would not have furnished a satisfactory basis for predictions as to the probable success of individual students in the college work of the first semester. For when we correlate the intelligence scores with the average college grades we get the following results:

	Correlation with college success	*P. E.*
Terman	.48	.05
Otis	.06	.07
Thurstone	.125	.065
Rogers	.34	.06
Average of the four tests	.45	.05

We undertook this study with a prejudice in favor of the tests and with the belief that the average scores secured from these four tests would furnish a much better basis for predicting college success than appears to have been the case. The coefficients in the case of the Terman and Rogers tests are good enough to be tantalizing. They also arouse the hope and justify the belief that it should soon be possible to improve group-intelligence testing to the point where we shall get correlations with college success that will run well above 0.60.

In this case the tests have done *something* with beautiful consistency, but that thing is not just exactly the thing that we desired to have done.

Hood College, Frederick, Maryland ELMER HOKE

A Revision of the Pressey Primer and Cross-Out Scales

Four years ago the writers developed two "brief group scales for measuring general ability for use in school surveys," the *Primer* Scale and the *Cross-out* Scale. The examinations were the result of a special effort at construction of scales which should be easy to give and score, inexpensive, and readily handled. They involved many radical departures from current practices in test building. In consequence, they were offered to the educational public with some misgivings as to the reception they might be accorded. Their wide-spread use, since that time, would indicate that they filled a real need. However, numerous short-comings have also appeared. It has therefore seemed advisable to attempt a revision of the series—a revision which this communication aims briefly to summarize.

The revision of the Primer scale has aimed primarily at better adaptation to work with first-grade children. One test of the first scale (the form or block test) has been entirely eliminated, as too difficult to explain to young children. The revised blank shows on the first page a directions test. This serves not merely as a test; it also accustoms the children to the procedure of the examination—habituates them to moving across the page from left to right (something many first-grade children have not yet learned), and so on. After taking the first test the children simply turn over the test folder instead of opening it; the second test is a revised pictorial absurdities test. The blank is next opened to an improved pictorial selection test showing four objects to each picture instead of three, in order to cut down the possibility of chance success. A revised and improved "dot pattern" test comes last. That the scale has in fact been made much more applicable to the first grade may be appreciated from the fact that, in a large school recently surveyed by the writers, no I-B child scored below 8 points. The scale is even more easily given and scored than the original Primer scale. More extensive norms than are available for the first form are now available for this revision.

The first Primer scale was found not entirely satisfactory for the third grade; third grade retards scored high, as a result merely of maturity. So the new Primer scale has been designed for use in the first two grades only. And an "Intermediate Classification Test" has been devised, to cover grades II-VI inclusive. The scale can perhaps hardly be called a "revision" of the Cross-out scale. But it is the direct result of the experience gained with that scale. It is similarly compact, and even more easy to use. On the first page of the small four-page folder appear the directions for all four tests of the scale, with examples from each. On the succeeding three pages are the 96 questions of the test proper, the four tests being arranged in cycle-omnibus form. The directions for the entire scale are thus given once for all at the beginning of the examination period; after this the examiner has nothing to do but collect the blanks when the time limit is reached.

The "Senior Classification Test" for Grades VII-XII is similar in organization to the intermediate scale; but the questions are, of course, of a more advanced and difficult nature. By thus covering with two scales the range of grades previously covered by the Cross-out scale it has been possible to

secure a much better differentiation; the grade over-lapping, for instance, is much less.

It is believed that this series of scales will prove of distinct value as brief instruments, for measuring general ability, for use in school surveys, and in other projects where an inexpensive, easily used, compact examination is desired. It may be added in closing that complementary examinations for each scale are contemplated, after the fashion of the Scale A and Scale B of the National Tests; when more reliable ratings on individual cases are desired it will then be possible to give the second form on a succeeding day. The total rating will then be much freer from the influence of chance circumstances than the findings from a single long examination.

S. L. PRESSEY
Ohio State University L. C. PRESSEY

Superintendent W. W. Lewton of Cicero, Illinois, informs us that his school system has been working on the "project" for the past three years. They have experimented, weighed results, and studied all the literature they could find on the subject. He feels that they are now seeing light both as to procedure and content.

The following words from a letter are self-explanatory: "We have found some big hurdles that must be cleared. First, the supervisors must be extremely well informed and they must be expert salesmen. Second, the proposition has to be sold to the teacher because the procedure is such a radical about-face in schoolroom activity that the 'informational' type of teacher, as well as the one deficient in leadership and initiative, will never be able to obtain results. Third, printed material, outlines, and suggestions are not extant. Fourth, our effort is pioneer work. We find it so because the 'problem project' is such a catchy phrase that many people have looked upon it as a panacea for all school ills."

We hope that Superintendent Lewton will organize the material which he has been gathering in such a way that it may be presented to the educational public.

Some pertinent sentences in the Bulletin on the Measurement of Ability in Spelling by the Bureau of Reference, Research, and Statistics of New York City are worth passing on for the thoughtful consideration of our readers. Read them carefully, and consider the extent to which you are applying the truths stated therein in your own school.

Improve your instruction by means of standard tests and measurements. The setting up of educational standards has resulted in the elimination of useless subject-matter. Economy demands that no more time be given to a subject than is necessary for the pupils to attain a satisfactory standard of achievement. Such standards are provided by educational measurements. Simply to know a pupil is below standard in ability is of little value to the teacher, because spelling ability is specific and not general. It is important that the teacher think of tests and scales as instruments which will enable her to make her instructional efforts more effective. Drill to be efficient, must be made individual in character.

Educational Research Association

(Frank N. Freeman, *Secretary and Editor*)

An Analytical Study of the Results of the Woody Arithmetic Test in West Allis, Wisconsin

Communicated by Mr. T. L. Torgerson,
Department of Educational Measurements

The Woody Arithmetic Scales were given to 1282 pupils in grades III to VII. The tabulation of the scores indicated that the pupils were up to or above the norm. The feature of the test which is of most interest to us, however, is the use which was made of the results to determine the frequency of error and the causes of these errors. The report (a part of which is here summarized) contains a considerable list of errors for each operation and the frequency with which these errors occurred. Space will not permit the printing of the full list. Table I, however, gives a summary of data for the five most frequent errors in each operation:

TABLE I. THE FIVE MOST PREVALENT TYPES OF ERROR IN EACH OPERATION

Error Number	Problem Number	Incorrect answers obtained	Frequency in each grade (percents)				
			III	IV	V	VI	VII
Addition							
1	35	9 ft. 20 in.			15.0	33.9	31.5
2	36	20 yr. 29 mo.			9.1	33.6	22.9
3	7	3	11.9	10.7	9.9	6.2	2.2
4	27	3.14		1.1	5.2	15.5	5.7
5	24	183762		2.1	5.0	4.2	10.5
Subtraction							
1	25	15 5/8			30.5	26.1	11.0
2	27	2 yd. 8 ft. 6 in.		4.2	26.5	24.5	13.0
3	35	2 2/8		6.4	21.0	21.2	18.6
4	14	35	20.3	14.0	6.6	2.4	
5	14	1 yd. 9 ft. 3 in.	21.7	13.2	4.6	2.8	
6	26	30		3.2	28.6	33.6	13.5
Multiplication							
1	34	1			10.1	26.8	11.7
2	35	24450 3/4			10.1	20.0	1.3
3	21	48 1/2		4.2	16.9	9.6	2.0
4	19	384	4.2	3.9	14.9	2.4	4.8
5	22	40 3/4		4.2	13.2	12.0	1.3

Division

1	35	71 21 531			.5	20.4	26.0
2	31	$\frac{3}{5}$			1.7	26.1	16.9
3	26	23		.7	8.9	26.1	.9
4	22	675		5.6	24.5	19.6	5.8
5	33	72 12 52			2.6	22.0	13.5

TABLE II. CAUSES OF ERRORS

Error Number	Cause
	Addition
1.	Failure to reduce to higher denomination in denominate numbers.
2.	The same as 1.
3.	Multiplication for addition.
4.	In fractions, adding the numerators for a new numerator and the denominators for a new denominator.
5.	Failure to use the decimal point.
	Subtraction
1.	Bringing down the fraction from a mixed number subtrahend.
2.	Ignorance of table of linear measure.
3.	Failure to reduce answer to lowest terms.
4.	Failure to pay back.
5.	Failure to borrow.
6.	Ignorance of table of linear measure.
	Multiplication
1.	Addition for multiplication.
2.	Failure to multiply by the fractional part of a mixed number.
3.	In mixed numbers, multiplying integers and adding fractions.
4.	Decimal point placed wrong.
5.	Failure to multiply by the fractional part of a mixed number.
	Division
1.	$a \times 0 = a$.
2.	Failure to place decimal point.
3.	Errors whose cause could not be discovered.
4.	The same as 3.
5.	$a \times 0 = a$.

From the data which are given elsewhere in the report it is possible to determine the causes of these most frequent errors. They are given in Table II. A number of very interesting facts appear from this table. The first is that, even though the number of errors here represented is small, we find repetitions of the same cause. The same cause appears more than once in a single operation and is also responsible for the mistakes in different operations. For example, the repetition of the failure to place the decimal point properly is found in several groups. Another interesting fact is that the errors in general appear to be rather trivial. They do not seem to depend

in some cases so much upon the failure to understand as upon mere carelessness. A further elaboration of this type of study would prove very valuable as a means of correcting or anticipating the errors which pupils make in their arithmetic.

The Result of a State-Wide Survey of Silent Reading in Iowa

Communicated by H. A. Greene,
Extension Division, State University of Iowa

In the survey which is here reported, the Monroe Silent Reading Test was given to a large number of schools and systems in Iowa, and the results were tabulated by the Extension Division. In all, results were obtained from 232 school systems, some situated in rural districts, others in cities. The distribution of the contributing school systems according to size is given in Table I.

TABLE I. CONTRIBUTING SCHOOL SYSTEMS[1]

Group	Population	Number of Cities
I	Less than 1,000	55
II	1,000 to 1,999	45
III	2,000 to 2,999	12
IV	3,000 to 4,999	11
V	5,000 to 9,999	12
VI	10,000 and above	12

[1]In addition to the city systems listed in this table, 85 consolidated schools also contributed.

Approximately 110,000 copies of the Monroe Silent Reading tests, Form 1, were shipped to the cities cooperating. From the 147 cities and 85 consolidated school finally reporting, records for 75,276 pupils in grades III to XII inclusive were received. Table II shows the distribution of these pupils by grades and by size of city.

TABLE II. NUMBER OF SCHOOLS AND PUPILS

Grade	Consolidated	Group I	Group II	Group III	Group IV	Group V	Group VI	Total
III	1,700	696	950	506	456	1,100	4,291	9,699
IV	1,571	687	1,076	449	550	1,087	4,134	9,554
V	1,549	727	1,142	421	645	1,085	4,020	9,589
VI	1,566	752	1,081	440	597	1,148	4,047	9,631
VII	1,521	765	1,173	447	586	1,014	3,607	9,113
VIII	1,532	721	1,252	420	554	875	3,300	8,654
IX	1,005	706	853	233	286	876	2,726	6,685
X	799	536	762	112	225	666	1,798	4,908
XI	627	429	669	125	249	622	1,425	4,149
XII	506	321	575	109	179	531	1,073	3,294
Total	12,376	6,340	9,533	3,272	4,327	9,004	30,424	75,276

The comprehension medians for each grade and group were computed from a composite distribution for each group of cities. That is, the scores made by the 1700 third-grade children from the 85 consolidated schools involved were distributed in a composite frequency table. From this composite table the median was calculated.

In the case of the rate medians lack of time prevented the making of similar composite distributions. Accordingly, in this case a weighted mean of the rate medians for each city group was found. This was checked sufficiently to show that the mean was fairly reliable and did not vary noticeably from a median calculated from a composite distribution. Evidence recently submitted with regard to the significance of Monroe rate scores leads to the conclusion that these rate scores are sufficiently accurate for ordinary purposes.

A comparison of the relative level of achievement of each grade of the different city groups with Monroe mid-year norms is shown in Table III. This table shows the percent that each grade median is of the mid-year norm. It is read as follows: the median score for third-grade pupils in the 85 consolidated schools is 67.5 percent of the Monroe norm for the middle of the year; the fourth-grade median for consolidated schools is 80 percent of the mid-year standard, etc.

TABLE III. PERCENT THAT GRADE MEDIANS ARE OF MONROE MID-YEAR NORMS

Grade	Consolidated Schools	Group I	Group II	Group III	Group IV	Group V	Group VI	Total	Monroe Mid-year Norms
					Comprehension				
III	67.5	67.5	65.0	106.0	92.5	83.8	128.0	93.0	6.8
IV	80.0	78.5	93.5	101.0	102.0	85.8	112.0	97.5	12.7
V	75.3	82.5	83.5	95.5	87.5	92.0	104.0	94.0	17.8
VI	92.5	92.9	114.0	121.0	109.0	103.0	113.0	107.0	18.5
VII	88.0	88.0	102.0	109.0	106.0	98.0	107.0	101.0	22.8
VIII	86.0	84.5	86.5	104.0	112.0	99.8	103.0	96.5	26.0
IX	97.8	98.0	104.0	109.0	127.0	110.0	123.0	113.0	23.0
X	104.0	103.0	102.0	108.0	121.0	113.0	121.0	113.0	25.4
XI	104.0	104.0	112.0	104.5	123.0	124.0	123.0	116.0	27.2
XII	105.0	105.0	112.0	97.0	132.0	114.0	120.0	116.0	30.0
					Rate				
III	79.5	69.5	85.5	100.0	96.5	92.5	115.0	91.1	52.0
IV	97.0	87.5	91.5	101.0	100.0	85.5	111.0	96.5	70.0
V	90.5	92.0	93.0	108.0	110.0	101.0	107.0	101.0	87.0
VI	96.0	92.0	95.0	106.0	109.0	103.0	108.0	101.0	90.0
VII	92.7	92.1	94.1	109.0	110.0	98.6	105.0	100.0	106.0
VIII	88.3	92.0	103.0	104.0	115.0	99.0	109.0	101.0	83.0
IX	95.0	95.0	103.0	100.0	108.0	98.0	114.0	102.0	85.0
X	98.2	94.3	99.0	106.0	119.0	103.0	111.0	104.0	85.0
XI	98.5	93.5	100.0	104.0	111.0	105.0	106.0	103.0	90.0
XII	95.8	94.0	98.0	98.5	126.0	98.0	104.0	101.0	96.0

The average scores or median scores made by the various groups of pupils may be obtained by multiplying the Monroe norms in the last column by the percents given in the main body of the table.

From Table II it is clear that the Iowa results fall below the Monroe mid-year comprehension standards in grades III, IV, V, and VIII. This situation can scarcely be considered serious except in the case of the lower grades in certain of the small grade schools. It is significant that in practically all cases the composite results for the high-school grades lie well above the mid-year standards established by Monroe for these grades. It is also noticeable from these data that the Iowa results move forward uniformly from grade V to grade VI instead of showing the typical break in the curve as revealed by Monroe's results. As a matter of fact, from the Iowa results it appears that Monroe's norms for the sixth and ninth grades are considerably lower in proportion than those for the fifth and eighth grades respectively. It also appears that so far as rate of reading is measured by the Monroe test the schools tested in this survey are approximately up to standard achievement for the middle of the year.

The foregoing table offers some indication as to the relative level of silent-reading ability in the different classes of cities. It will be noticed that for cities of Group III the composite median for grades V and XII are the only ones which fall below the Monroe norms. In the case of Group VI no composite median falls below the norm. A consideration of these results seem to justify the conclusion that due to certain factors not identified in this report a higher quality of silent-reading comprehension has been developed in the larger school system.

The foregoing data indicate the level of achievement of classes in various types of city schools at the mid-year. In addition to these data, tables not reproduced here, have been constructed to show the large variability among the scores in individual grades. It is planned to give the test again at the end of the second semester. This will indicate the growth which has then taken place during a half year.

Request for Suggestion on Administrative Problems

Assistant Superintendent Joseph P. O'Hern of Rochester, New York, wishes to secure information or suggestions on the following topics:

1. Promotion.—Based on class work, examination, or both; trial promotions and, if allowed, for what length of period; promotion based on course of study arranged for maximum, average, and minimum essentials; responsibility for non-promotion placed on receiving or sending teacher.

2. Per capita cost.—What items are figured into the cost, and what divisors are used.

3. The average number of pupils per teacher and for various types of schools.

4. Salary schedules for different types of service.